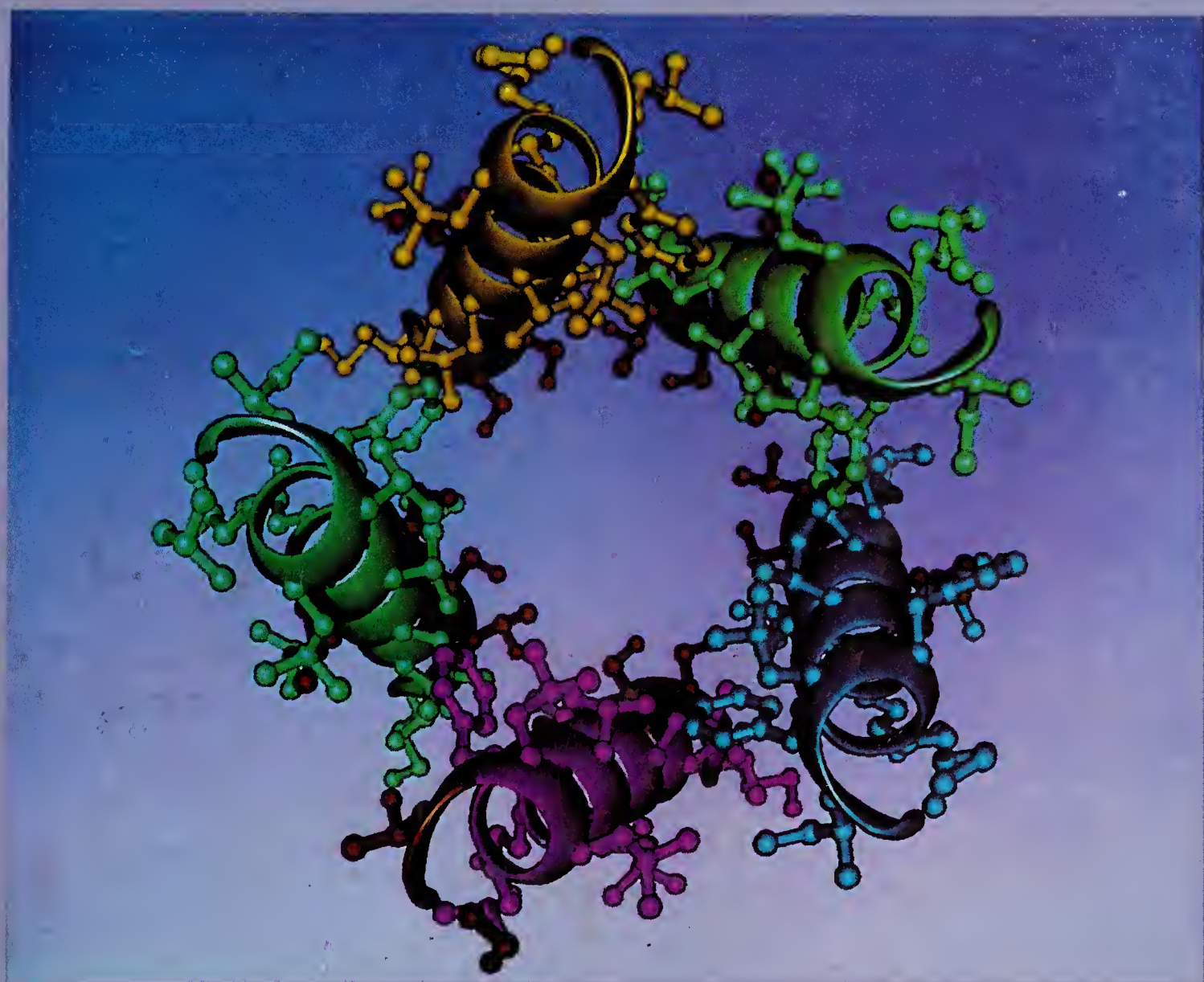


MOLECULAR CONFORMATION AND BIOLOGICAL INTERACTIONS

Editors

P. Balaram and S. Ramaseshan



G. N. RAMACHANDRAN FESTSCHRIFT

**INDIAN ACADEMY OF SCIENCES
BANGALORE**

Advances in molecular biology, crystallography, spectroscopy, chemical synthesis and computer graphics have been central to the explosive development of the area of structure–function relationships for biomolecules. This tribute to G. N. Ramachandran, one of the pioneers in the area of protein structure and conformational analysis, brings together over forty state-of-the-art articles in areas ranging over protein structure and function, modelling and design of novel proteins, membrane channels, nucleic acid structures and drug interactions, peptide conformations and synthetic peptides as biological models, drug design and molecular recognition.

Topics covered include: models for predicting protein structures; side-chain packing; looping and motifs in proteins; crystallographic studies on cytochrome P-450, insulin and carbonic anhydrase; flavoenzymes as electron transfer models; glycolytic enzymes as drug targets; protein hydration; crystallins; sickle cell haemoglobin polymerization; spermine:nucleic acid complexes;

MOLECULAR CONFORMATION **and** **BIOLOGICAL INTERACTIONS**

G. N. Ramachandran Festschrift



Digitized by the Internet Archive
in 2018

MOLECULAR CONFORMATION and BIOLOGICAL INTERACTIONS

G. N. Ramachandran Festschrift

Editors

P. BALARAM

Indian Institute of Science, Bangalore

S. RAMASESHAN

Raman Research Institute, Bangalore



1 9 9 1

INDIAN ACADEMY OF SCIENCES

Bangalore

Published by

Indian Academy of Sciences
CV Raman Avenue
P.B. No. 8005, Sadashivanagar
Bangalore 560 080, INDIA

Tel: (0812)342546
Telex: 0845-2178 ACAD IN
Telefax: 91-812-346094

ISBN 81-85324-12-3

Copyright © 1991 by the Indian Academy of Sciences

Typeset and printed by Thomson Press (India) Ltd., Faridabad

CONTENTS

Foreword	xi
Analysis and Prediction of Protein Structures	
Towards Understanding How Side Chains Pack in Proteins <i>Juswinder Singh, Lilian C. Nandi and Janet M. Thornton</i>	1
Long and Medium-Sized Irregular Regions in Proteins as Combinations of Small Standard Structures <i>A. V. Efimov</i>	19
Simplified Models for Understanding and Predicting Protein Structure <i>John M. Troyer and Fred E. Cohen</i>	31
Conformations for Strand Entry into Parallel β -Sheets <i>Cyrus Chothia and Arthur M. Lesk</i>	49
Analysis of Short Loops Connecting Secondary Structural Elements in Proteins <i>N. Srinivasan, R. Sowdhamini, C. Ramakrishnan and P. Balaram</i>	59
Protein Structure and Function	
Structural Studies on the Mechanism of O ₂ /Peroxide Activation and Electron Transfer in Peroxidase and Cytochrome P-450 <i>Thomas L. Poulos</i>	75
Conformational Changes at the B Chain N Terminus in the Insulin Hexamer <i>G. G. Dodson, R. E. Hubbard and B. Xiao</i>	97
Protein Crystallography and Enzyme Catalysis <i>Georg E. Schulz</i>	115
Structural Perspectives in the Function of Carbonic Anhydrase <i>K. K. Kannan</i>	131
Variability in Protein Hydration and its Structural Consequences – A Novel Approach Involving Water-Mediated Transformations <i>M. Vijayan and R. Kodandapani</i>	149

Structure of Bowman-Birk Protease Inhibitor and its Complex with Trypsin <i>Atsuo Suzuki, Yasuo Tsunogae, Takashi Yamane and Tamaichi Ashida</i>	171
Complex Flavoenzymes – Models for Electron Transfer Complexes <i>F. Scott Mathews</i>	187
Crystallographic Investigations of Glycolytic Enzymes from <i>Trypanosoma brucei</i> : Potential Starting Points for the Design of New Sleeping Sickness Drugs <i>Wim G. J. Hol, Fred M. D. Vellieux, Christophe L. M. J. Verlinde, Rik K. Wierenga, Martin E. M. Noble and Randy J. Read</i>	215
Protein Interaction, Folding and Energetics Studied by Hydrogen Exchange Labeling <i>S. Walter Englander, Leland Mayne, Russell E. McKinnie, Yvonne Paterson and Joan J. Englander</i>	245
Polymerization of Hemoglobin S Amidated at the Contact Residue Glu 43(β) <i>A. Seetharama Acharya, Lakshmi Khandke, Brian T. Chait and M. Janardhan Rao</i>	269
Crystallin Structure and Eye Lens Function <i>Manni Luthra, C. Sivakama Sundari, P. Guptasarma and D. Balasubramanian</i>	281
Nucleic Acids and Drug Interactions	
Molecular Structures of Spermine: Nucleic Acid Complexes <i>Loren Dean Williams, Christin A. Frederick, Reinhard V. Gessner and Alexander Rich</i>	295
Stereochemical Studies on DNA and DNA–Drug Interactions <i>S. Mohan and N. Yathindra</i>	311
Solvent Accessibility of Nucleic Acids <i>P. K. Ponnuswamy</i>	335
DNA Bending: A Natural Consequence of Base Sequence Dependent Variability <i>Manju Bansal, Dhananjay Bhattacharyya and Debasisa Mohanty</i>	347
Molecular Recognition	
Molecular Recognition in Peptides and Proteins <i>Andrew J. Doig and Dudley H. Williams</i>	363

Molecular Recognition in Aspartic Proteases	385
<i>H. Mattras, M. Bianchi, J. A. Fehrentz, R. A. Boigegrain, L. Chiche, M. A. Coletti-Previero and B. Castro</i>	
Dynamic Aspects of Protein Specificity	401
<i>G. C. K. Roberts</i>	
Protein–Carbohydrate Interactions – A Theoretical Approach	411
<i>V. S. R. Rao</i>	
Peptide Conformation and Design	
The Ramachandran Plot and Beyond: Conformational and Topographical Considerations in the Design of Peptides and Proteins	429
<i>Victor J. Hruby and Gregory V. Nikiforovich</i>	
The Design of Novel Proteins	447
<i>Bernd Gutte and Stephan Klauser</i>	
Template Assembled Synthetic Proteins (TASP): A Chemical Approach to the Construction of New Proteins	457
<i>Robert I. Carey and Manfred Mutter</i>	
Peptide Models for Reverse Turns. The Role of Asparagine in the <i>i</i> Position of a β -turn	469
<i>Josep Rizo, Madan M. Dhingra and Lila M. Gierasch</i>	
The 3_{10} - and α -Helical Conformations in Peptides	497
<i>Ettore Benedetti, Benedetto Di Blasio, Vincenzo Pavone, Carlo Pedone, Antonello Santini, Marco Crisma and Claudio Toniolo</i>	
Helical Transitions in Peptides	503
<i>John D. Clark, Edward E. Hodgkin and Garland R. Marshall</i>	
The Fully Extended Polypeptide Conformation	511
<i>Claudio Toniolo and Ettore Benedetti</i>	
A Guest–Host Approach to Oligodepsipeptides; Studies of Oligoglutamates and Depsiglutamates in Trifluoroethanol	523
<i>Y. V. Venkatachalapathi, R. Katakai and M. Goodman</i>	
Peptides as Models for Protein-Folding	537
<i>E. Subramanian</i>	

Peptides as Biological Models

- Thermally Driven Self-Assembly, Molecular Structuring and Entropic Mechanisms in Elastomeric Polypeptides 555

Dan W. Urry

- Use of Synthetic Peptides in the Structural and Functional Studies of Apolipoprotein A-1 585

Y. V. Venkatachalapathi, Jan G. Dohlman, Warren D. Blackburn, Kiran B. Gupta, Shamala K. Srinivas, Jere P. Segrest and G. M. Anantharamaiah

- Role of Basic Amino Acid Residues in the Amphipathic Helix: the Snorkel Hypothesis 597

Jere P. Segrest, Y. V. Venkatachalapathi, Shamala K. Srinivas, Kiran B. Gupta, Hans De Loof and G. M. Anantharamaiah

- 2D-NMR Conformational Analysis of Griselimycin, an Antituberculous Cyclodepsipeptide 611

Manh Thong Cung, Bernard Vitoux, Pascal Demange and Michel Marraud

Membrane Transport

- Structure and Function of Transmembrane Ion Channels 627

Robert M. Stroud

- Molecular Architecture by Using Cyclic Peptides for Ion-Transport Through Lipid Membrane 645

Yukio Imanishi and Shunsaku Kimura

- Signal Sequences: The Relevance of Their Conformations and Membrane Perturbing Ability in the Sorting of Secretory Proteins 661

R. Nagaraj

- Ionophores: Structure and Interaction in Relation to Transmembrane Ion-Transport 671

K. R. K. Easwaran

Viruses and Cellular Metabolism

- Structure and Assembly of Viruses 687

M. R. N. Murthy, C. N. Hiremath and H. S. Savithri

- Metabolism and Maturation Studies of Intact Spermatozoa Using Magnetic Resonance Techniques 703

Sudha Srivastava, Elvyra J. Fernandes, Umashashi C. Hegde, Ratna S. Phadke and Girjesh Govil

- Index 713



Foreword

The determination of the relationships between molecular structures and biological activity is central to contemporary approaches to the study of biological phenomena. Structural biology is a relatively young field with its first triumphs appearing in quick succession in the 1950s. Linus Pauling's postulation of the α -helix and β -sheet as central features of protein structures was followed by the Watson–Crick proposal of the double helical nature of DNA. Soon thereafter, the triple helical structure of collagen, the ubiquitous connective tissue protein, was recognised by Ramachandran and Kartha working at Madras in 1954. The late 1950s and early 1960s also saw the determination of the first three-dimensional structures of globular proteins by X-ray diffraction methods, by Kendrew and Perutz. Very soon the field exploded into activity with protein, peptide and nucleotide structures being determined by an increasing number of groups, worldwide. In the 1980s crystallography was joined by powerful NMR methods which permit three-dimensional structure determination in solution. The enormous scope of experimental methods has been enhanced by the growing range of computational techniques, supported by the ever increasing power of modern computers. The field of structural biology thus appears poised today, to attack the many formidable problems that remain to be solved in rationalising conformation–function relationships, for a wide variety of biochemical problems.

A landmark contribution in the understanding of protein structures in particular, and biopolymer structures in general, was the introduction of hard sphere criteria as a stereochemical filter in limiting the range of accessible molecular conformations. This approach to biomolecular conformation was pioneered by G. N. Ramachandran at Madras in the 1960s. The Ramachandran map is the most common representation of backbone conformations of individual residues in proteins. This Festschrift is a tribute to G. N. Ramachandran, a scientist who has made seminal contributions to diverse areas of X-ray crystallography, conformational analysis of biomolecules, fibrous protein structures and three-dimensional image reconstruction. Ramachandran has been the driving force behind the development of crystallography and molecular biophysics in India. In the course of his career he established two major centres, at the University of Madras and the Indian Institute of Science, Bangalore. The scientific community in India has been greatly stimulated by Ramachandran's academic contributions. This Festschrift produced by the Indian Academy of Sciences is an acknowledgement of his dominant role in shaping biological research in India.

We are particularly grateful to authors from several countries for providing a current perspective of their research for inclusion in this volume. We thank Vinod Mishra, N. Srinivasan and R. Gurunath for help in various stages of preparation of this Festschrift. The production of this book owes a great deal to the work done by the staff of the Indian Academy of Sciences, particularly K. Shashikala, Chandrika Ramesh and G. Madhavan.

P. Balaram
S. Ramaseshan

Towards Understanding How Side Chains Pack in Proteins

Juswinder Singh, Lilian C. Nandi and Janet M. Thornton

Biomolecular Structure and Modelling Unit, Department of Biochemistry
and Molecular Biology, University College London, Gower Street,
London WC1E 6BT, UK

Introduction

Central to the rational mutation of a residue in a protein, or the successful docking of a peptide ligand into a binding site is an understanding of how amino acid side chains pack against each other. Knowledge of the forces holding proteins together has been gained by analysis of the wealth of structural information obtained from X-ray crystallographic studies of small molecules and macromolecules. Over 400 protein structures have been determined, with a significant number at high resolution (1). Attention has primarily been focused on understanding the backbone conformation, and identifying the possible motifs adopted. Little is known about the packing of the side chains, and yet it is the interaction between these groups which determines the structure, function and stability of proteins.

Proteins are polymers of 20 basic units called amino acids. These 20 groups consist of a standard peptide part, and a variable side chain. Up to 400 pair-types are possible (20×20 side chain pairs). Several crystallographic analyses have been carried out on side chain packing. These have included both statistical analysis of the geometries, and use of local coordinate systems to transform interacting pairs into common orientations, to compare packing arrangements (2–5). Studies of aromatic–aromatic interactions have shown that the rings often pack with an edge-face arrangement, but are never seen to stack with their ring planes parallel and fully overlapped. Ion pairs between Arg and Asp, Glu side chains within proteins show a 16:1 preference to form twin hydrogen bonds with the $N\epsilon$, NH_2 rather than the NH_1 and NH_2 (6). These results are therefore valuable in understanding the constraints on side chain packing and the forces of interaction between protein groups. However they represent a small fraction of the data available. Due to the importance of the data we have extended our approach to calculate the geometries of all the 400 possible types of side chain–side chain interactions in 52 high resolution protein structures, and also the 380 possible side chain–atom distributions (7). Here we describe the database, and give specific examples of side chain–side chain, and side chain–atom packing in proteins. We also present results for using PAIRS distributions to model inhibitor side chains to their receptors. Finally, we discuss a graphics interface to the database, and a dictionary of the data, and how they may be used.

Methods

The database of side chain interactions was derived from 52 protein crystal structures deposited in the Brookhaven databank (1). Since the geometry of the side chains were being considered, only high resolution data were used (resolution ≤ 2.0 Å). Using a suite of FORTRAN routines – SIRIUS, we have determined the pairwise geometry between all 400 pair-types, and the 380 types of side chain–atom interaction (7). The following procedure was used (figure 1):

1) For each of the 52 proteins, the closest interatomic distance was calculated between all pair type residues. For example, if Arg–Asp interactions were being considered, all Arg and Asp residues would be extracted, and checked for close contacts. If the closest distance R was less than the summation of the respective van der Waals radii of the atoms, plus 1.0 Å, they were regarded as interacting.

2) Interacting pairs were then transformed into a common reference frame. This involves fitting the first residue of the pair-type (e.g. Arg of an Arg–Asp pair) onto a reference group (reference Arg; figure 2) constructed from ideal bond lengths and angles, and then applying the resulting transformation matrix to the Asp. This procedure when repeated for all interacting pairs gives a geometric distribution of

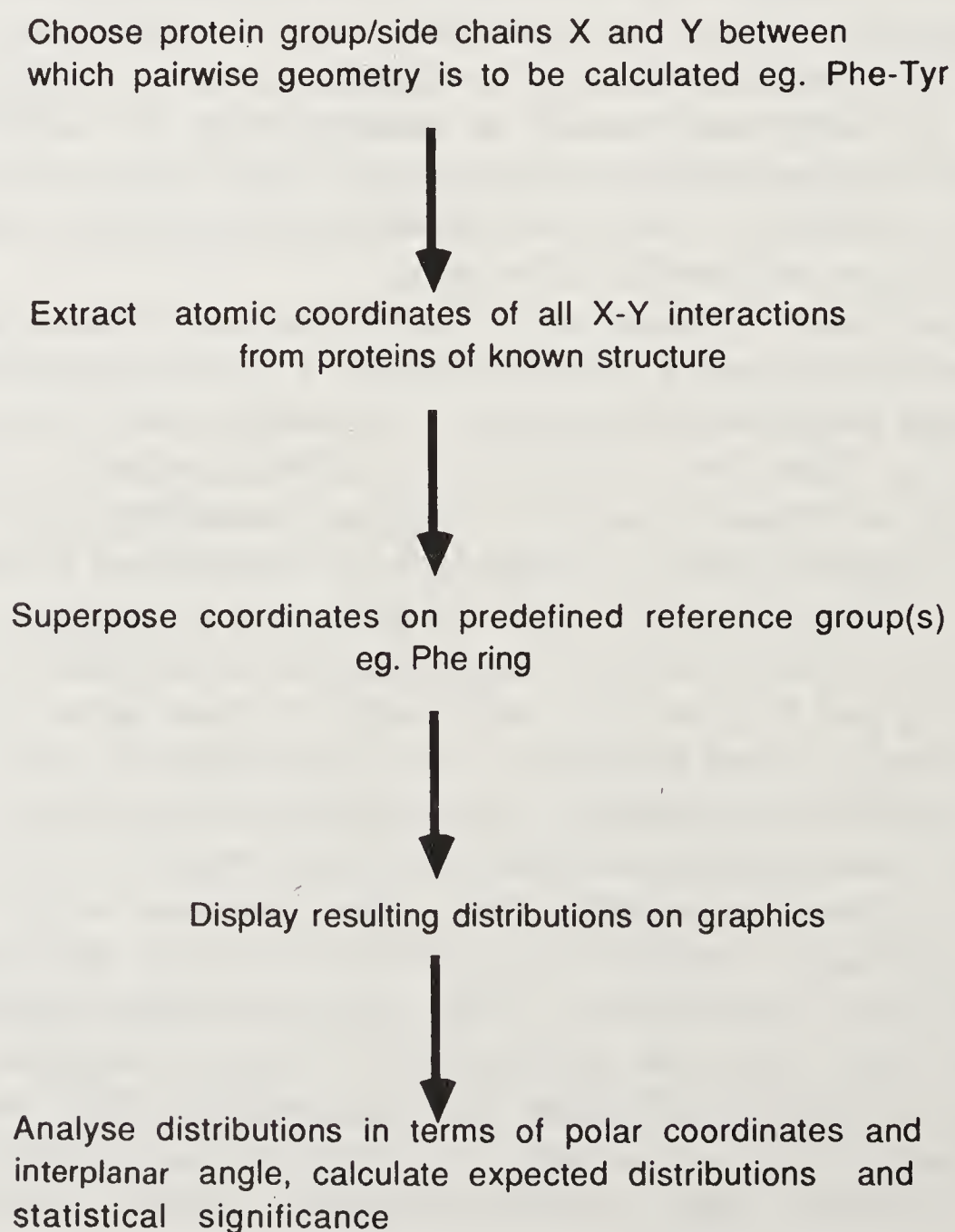


Figure 1. Flowchart of method employed in calculating side chain–side chain interactions.

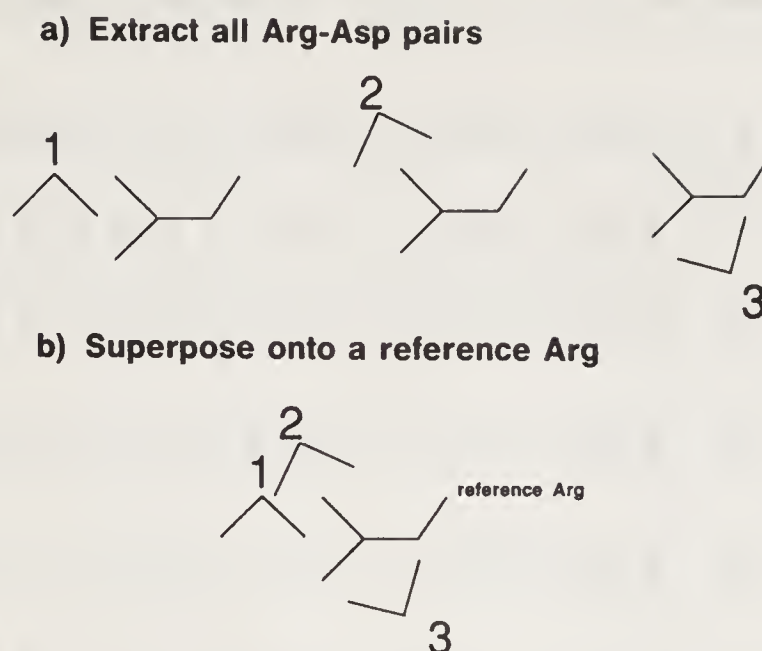


Figure 2. Procedure used for generating a PAIRS distribution. This involves superposing interacting PAIRS into a common reference frame. For an Arg-Asp PAIRS distribution, the first step is to extract the Arg-Asp pair from the protein database, then superpose it onto a reference Arg group located at the origin of the coordinate system. After repeating this procedure for all Arg-Asp pairs, a three-dimensional distribution of Asp groups is obtained around a reference Arg.

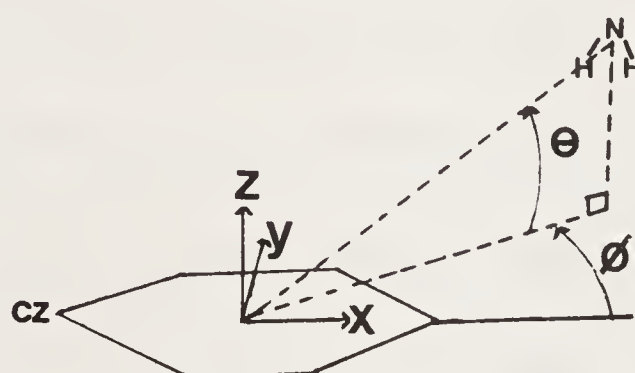


Figure 3. Definition of the spherical polar angles θ and ϕ for a Phe-NH₂ pair. θ defines the position of the NH₂ group above the plane of the Phe, ϕ defines where the NH₂ group lies in the x - y plane of the Phe.

the pair interaction in three-dimensional space. Since all interacting pairs now share the same coordinate system, their geometries can be compared.

3) The distributions were examined on the graphics, and also in terms of a distance of closest approach R , and three angles used to define their geometry. The orientation between the interacting pairs of side chains was found by calculating the dihedral angle P between the planes of the two side chains. To define the spatial relationship between the interacting side chains, the spherical polar angles, θ and ϕ were used (figure 3). Random distributions for P , R , θ and ϕ were calculated by assuming all volume to be equally filled, to assess the significance of the observed distributions.

Results

Database of Side Chain Packing in Proteins

Table 1 shows the observed frequencies of side chain interactions in our 52 protein dataset. In total there are 10,202 side chains in the dataset, and 34,034 interacting

Table 1. Observed frequencies of interaction for the 400 side chain-side chain pair types.

	ALA	ASN	ASP	ARG	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
NRES*	858	456	587	307	281	335	483	984	242	518	761	554	172	403	453	808	693	170	410	727
ALA	96	63	61	48	31	58	40	166	41	177	222	56	36	140	68	61	118	78	112	194
ASN	63	94	78	48	29	65	66	108	38	50	83	84	20	53	68	63	95	36	86	82
ASP	61	78	50	105	21	34	47	113	57	60	63	131	17	60	53	94	106	31	74	49
ARG	48	48	105	44	25	53	94	75	32	60	94	23	22	64	51	49	75	36	60	62
CYS	31	29	21	25	234	26	9	83	25	64	84	16	30	64	46	28	34	25	58	79
GLN	58	65	34	53	26	38	35	83	22	60	92	40	22	51	63	57	75	33	68	79
GLU	40	66	47	94	9	35	46	47	55	66	74	124	22	55	48	74	84	36	50	72
GLY	166	108	113	75	83	83	47	242	57	158	214	83	48	111	111	73	204	83	103	206
HIS	41	38	57	32	25	22	55	57	74	41	69	32	22	56	45	41	60	29	54	59
ILE	177	50	60	60	64	60	66	158	41	376	436	75	87	243	83	72	130	105	143	333
LEU	222	83	63	94	84	92	74	214	69	436	622	90	126	355	135	102	155	138	188	481
LYS	56	84	131	23	16	40	124	83	32	75	90	28	24	57	52	46	91	32	89	84
MET	36	20	17	22	30	22	22	48	22	87	126	24	52	86	26	32	33	38	44	80
PHE	140	53	60	64	64	51	55	111	56	243	355	57	86	226	80	72	101	87	120	246
PRO	68	68	53	51	46	63	48	111	45	83	135	52	26	80	58	56	91	49	97	121
SER	61	63	94	49	28	57	74	73	41	72	102	46	32	72	56	76	135	30	56	97
THR	118	95	106	75	34	75	84	204	60	130	155	91	33	101	91	135	172	49	95	160
TRP	78	36	31	36	25	33	36	83	29	105	138	32	38	87	49	30	49	36	63	114
TYR	112	86	74	60	58	68	50	103	54	143	188	89	44	120	97	56	95	63	100	148
VAL	194	82	49	62	79	79	72	206	59	333	481	84	80	246	121	97	160	114	148	396

* NERS – number of residues in the database

Table 2. Ratios of observed to expected frequencies of side chain–side chain interactions.

	ALA	ASN	ASP	ARG	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
ALA	0.2	0.5	0.4	0.6	0.4	0.6	0.3	0.6	0.6	1.2	1.0	0.4	0.8	1.2	0.5	0.3	0.6	1.6	1.0	1.0
ASN	0.5	0.7	0.9	1.0	0.7	1.3	0.9	0.7	1.1	0.6	0.7	1.0	0.8	0.9	1.0	0.5	0.9	1.4	1.4	0.8
ASP	0.4	0.9	0.2	1.8	0.4	0.5	0.5	0.6	1.2	0.6	0.4	1.2	0.5	0.8	0.6	0.6	0.8	0.9	0.9	0.3
ARG	0.6	1.0	1.8	0.7	0.9	1.6	2.0	0.8	1.3	1.2	1.2	0.4	1.3	1.6	1.1	0.6	1.1	2.1	1.5	0.8
CYS	0.4	0.7	0.4	0.9	4.5	0.8	0.2	0.9	1.1	1.3	1.2	0.3	1.9	1.7	1.1	0.4	0.5	1.6	1.5	1.2
GLN	0.6	1.3	0.5	1.6	0.8	0.5	0.7	0.8	0.8	1.1	1.1	0.7	1.2	1.2	1.3	0.6	1.0	1.7	1.5	1.0
GLU	0.3	0.9	0.5	2.0	0.2	0.7	0.3	0.3	1.4	0.8	0.6	1.4	0.8	0.9	0.7	0.6	0.8	1.3	0.8	0.6
GLY	0.6	0.7	0.6	0.8	0.9	0.8	0.3	0.4	0.7	0.9	0.9	0.5	0.9	0.9	0.8	0.3	0.9	1.5	0.8	0.9
HIS	0.6	1.1	1.2	1.3	1.1	0.8	1.4	0.7	1.9	1.0	1.1	0.7	1.6	1.8	1.3	0.6	1.1	2.2	1.7	1.0
ILE	1.2	0.6	0.6	1.2	1.3	1.1	0.8	0.9	1.0	2.1	3.4	0.8	3.0	3.6	1.1	0.5	1.1	3.6	2.1	2.7
LEU	1.0	0.7	0.4	1.2	1.2	1.1	0.6	0.9	1.1	3.4	1.6	0.7	2.9	3.5	1.2	0.5	0.9	3.3	1.8	2.7
LYS	0.4	1.0	1.2	0.4	0.3	0.7	1.4	0.5	0.7	0.8	0.7	0.1	0.8	0.8	0.6	0.3	0.7	1.0	1.2	0.6
MET	0.8	0.8	0.5	1.3	1.9	1.2	0.8	0.9	1.6	3.0	2.9	0.8	2.7	3.7	1.0	0.7	0.8	3.8	1.9	2.0
PHE	1.2	0.9	0.8	1.6	1.7	1.2	0.9	0.9	1.8	3.6	3.5	0.8	3.7	2.1	1.3	0.7	1.1	4.0	2.2	2.6
PRO	0.5	1.0	0.6	1.1	1.1	1.3	0.7	0.8	1.3	1.1	1.2	0.6	1.0	1.3	0.4	0.5	0.9	2.0	1.6	1.1
SER	0.3	0.5	0.6	0.6	0.4	0.6	0.6	0.3	0.6	0.5	0.5	0.3	0.7	0.7	0.5	0.2	0.7	0.7	0.5	0.5
THR	0.6	0.9	0.8	1.1	0.5	1.0	0.8	0.9	1.1	1.1	0.9	0.7	0.8	1.1	0.9	0.7	0.5	1.3	1.0	1.0
TRP	1.6	1.4	0.9	2.1	1.6	1.7	1.3	1.5	2.2	3.6	3.3	1.0	3.8	4.0	2.0	0.7	1.3	1.9	2.7	2.8
TYR	1.0	1.4	0.9	1.5	1.5	1.5	0.8	0.8	1.7	2.1	1.8	1.2	1.9	2.2	1.6	0.5	1.0	2.7	0.9	1.5
VAL	1.0	0.8	0.3	0.8	1.2	1.0	0.6	0.9	1.0	2.7	2.7	0.6	2.0	2.6	1.1	0.5	1.0	2.8	1.5	1.1

The number of expected interactions is:

$$\text{exp} = \frac{\text{number of Res}_a \times \text{number of Res}_b}{\text{total number of residues} \times \text{total number of residues}} \times \text{Nint} \times \text{Norm}$$

where Res_a and Res_b are the number of residues of type a and b (e.g. Phe–Tyr). Nint is the total number of interactions, and Norm is a constant which is set to 2 if a and b are the same residue type, and 1 if they are not

pairs. For the majority of the pairs there are sufficient data to allow analysis. How can these data be used? Firstly, using these propensities we can infer the favourability of any pairwise interaction. This approach was used by Warne and Morgan (8) in 16 protein structures, and yielded much useful information on protein interactions. In table 2 we have calculated the ratios of the observed to expected frequencies of interaction, where the latter reflects the number of interactions expected from the occurrence of those residues in the sequences of the protein structures used to create the database. If the value in the table is greater than 1, then the interaction occurs more often than would be expected from its occurrence in the protein structures, suggesting favourability in the interaction. This is true for the ion-pair interactions (Arg–Asp; 1.8), and the aromatic groups (Phe–Phe; 2.0). However, if the value is less than 1, the interaction is seen less often than would be expected, and is therefore unfavourable. This is the case for Cys–Glu (0.2), and Lys–Lys pairs (0.1). Second, the database can be used to investigate packing preferences for pairs of amino acids. For each of the interacting pairs, the R , θ , ϕ , interplanar angle P , and transformed coordinates were stored. Using these data, we can compare observed geometries with those expected from equal occupation of space, and from this assess favourable and unfavourable geometries. To illustrate our approach we present His–His and Phe–Lys interactions as examples.

(i) *His–His*: Histidine contains a 5 membered heterocyclic aromatic side chain with two nitrogens. The observed to expected ratio is 1.9, therefore these groups like to interact. The distance profile in figure 4a, shows the closest interatomic distance between the side chain atoms of the histidines. The majority of the contacts occur between the side chain nitrogens, most of which lie between 2.8 Å–3.2 Å. This suggests the histidines have a preference to hydrogen-bond. This is supported by the stereo plot of the PAIRS distribution. Figure 5a which views the reference group along the z -axis, shows the distribution of histidines around the nitrogens of the reference group. The conformers form clear clusters around the two nitrogens of the histidine. There is also a preference for histidines to lie in the plane of the reference His, as expected for a hydrogen bonding interaction (figure 4b $\theta = 0^\circ$ – 10° ; figure 5a). Surprisingly, there are a larger number of histidines around the N ϵ 2 than the N δ 1, seen as a peak in the ϕ distribution at 140° – 160° (figure 4c). This could be explained in part by the preference for other groups to hydrogen-bond to the N δ 1. This is observed for the shorter hydrogen bonding amino acids such as Asp and Ser. These groups have limited flexibility, and if the interacting pairs are close together in the sequence they will be restricted to interacting via the N δ 1.

(ii) *Phe–Lys*: Phe is an aromatic side chain, consisting of an apolar six-membered ring, while Lys has a long carbon chain, capped with a positively charged amino group. Unlike the short contact distances of the polar His–His pairs, these groups tend to peak around the van der Waals distance of 3.4 to 3.8 Å (figure 6a). Figures 6c and 6d show two orthogonal views of the Phe–Lys distribution. There is no clustering of the Lys side chains in the x – y plane of the ring, unlike the view along the edge of the reference group. This shows a striking preference for the Lys side chains to avoid the edge of the aromatic ring, and a clear clustering for them above and below the plane of the ring. This is quantified in figure 6b, where the random distributions are shown. Close inspection of the PAIRS distributions shows no

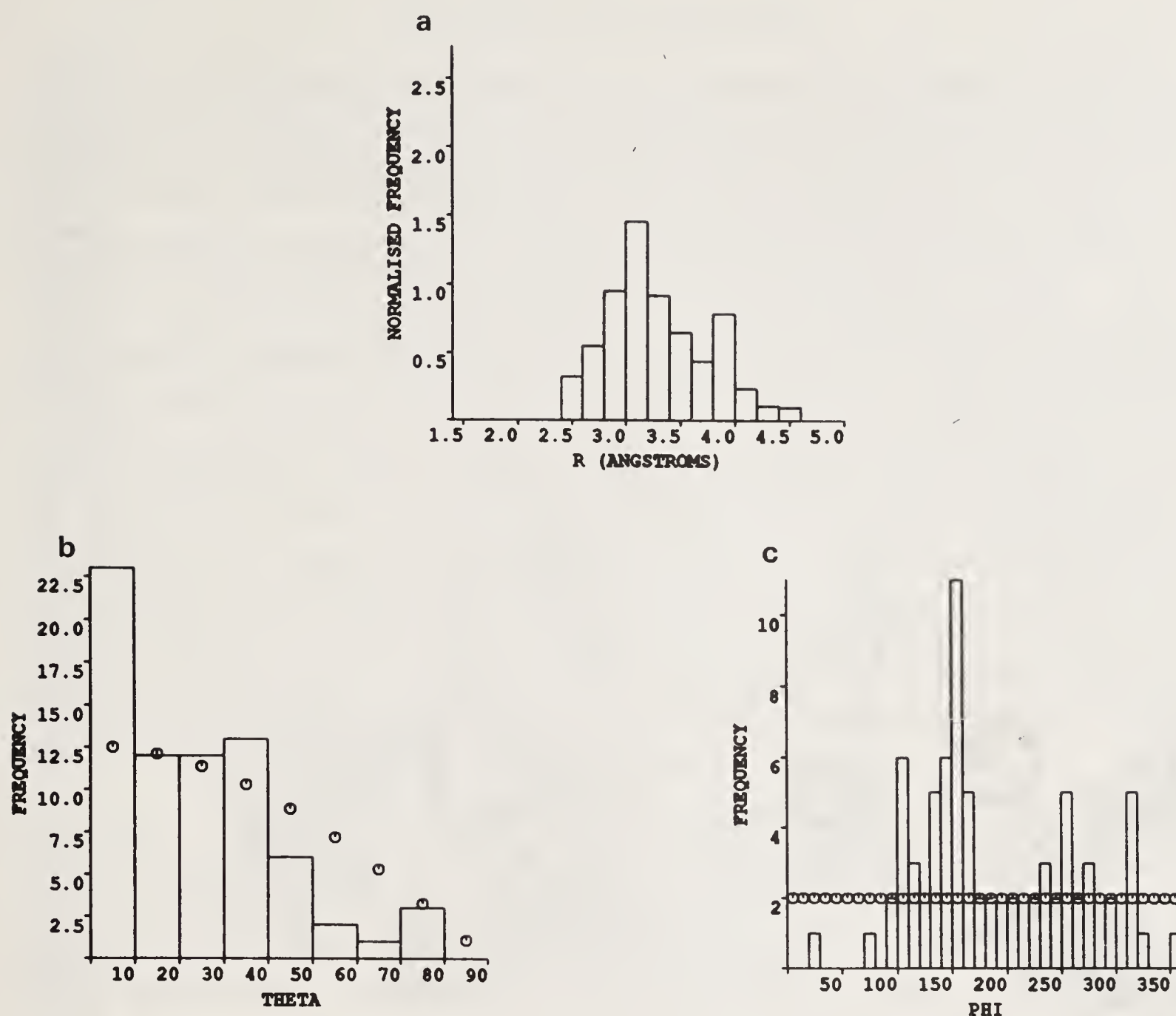


Figure 4. (a) Plot of the closest interatomic distance between the side chain atoms of the His groups in the database. The distribution has been normalised to take into account bias due to volume differences. Distribution peaks around 3.0–3.2 Å, as expected from a hydrogen-bond. (b) Plot of the angle θ . θ defines the position of the centre of the His ring above the plane of the reference His. Low values of θ show histidines pack in-plane, while high θ that His groups lie directly above the reference His. The random distribution is denoted by a circle. There are twice as many observations at low theta (0° to 10°) than expected by chance, suggesting a preference for oxygens to lie in the plane. (c) Plot of the angle phi. The random distribution is denoted by a circle. This indicates where histidines cluster in the x - y plane of the reference His. There is a clear peak in the distribution at 150° to 160° which corresponds to the cluster of histidines around the N ϵ 2 of the reference His.

preference for the amino groups to be directly above the face of the ring. The theta distribution confirms this, with a preference for lysines to lie above the edge, and below the face of the Phe ring ($\theta = 30^\circ$ – 70°). The avoidance of the plane of the ring could be accounted for by the unfavourable interactions of the partial positive charge on the CH ring edge of the aromatic ring with the positively charged amino group. The amino groups avoidance of the ring face is contrary to suggestions that a hydrogen bond exists between the nitrogen's hydrogens and the π electron system of the aromatic ring (9). These results suggest that amino- π interactions are unfavourable.

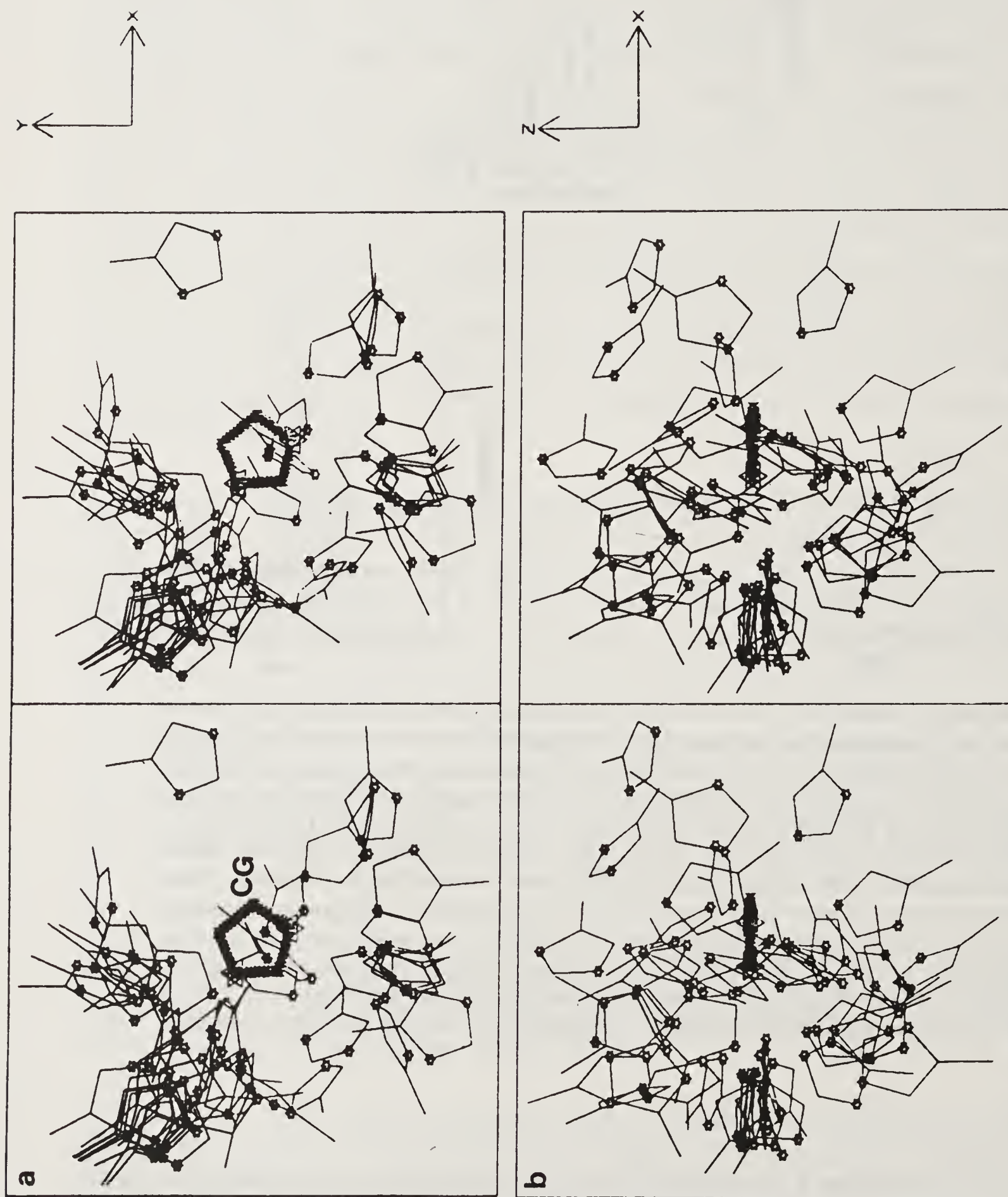


Figure 5. Stereo plots of His-His PAIRS distribution. (a) View down the z-axis towards the face of the reference His group. There is a large cluster of histidines around the Nε2 of the reference His. (b) View of the PAIRS distribution along the y-axis towards the edge of the reference group.

Side Chain–Atom Interactions

There are 19 different types of atoms in the 20 amino acids (10) and it is important to see whether preferred geometries of side chains can be rationalised at the atomic level. There are 19 different atomic distributions to be evaluated around each of the 20 side chains (380 PAIRS distributions). From this approach we can explore the complete atomic environment of a side chain, and assess similarities and differences between the atom types. Here we consider the atomic environment of Arg.

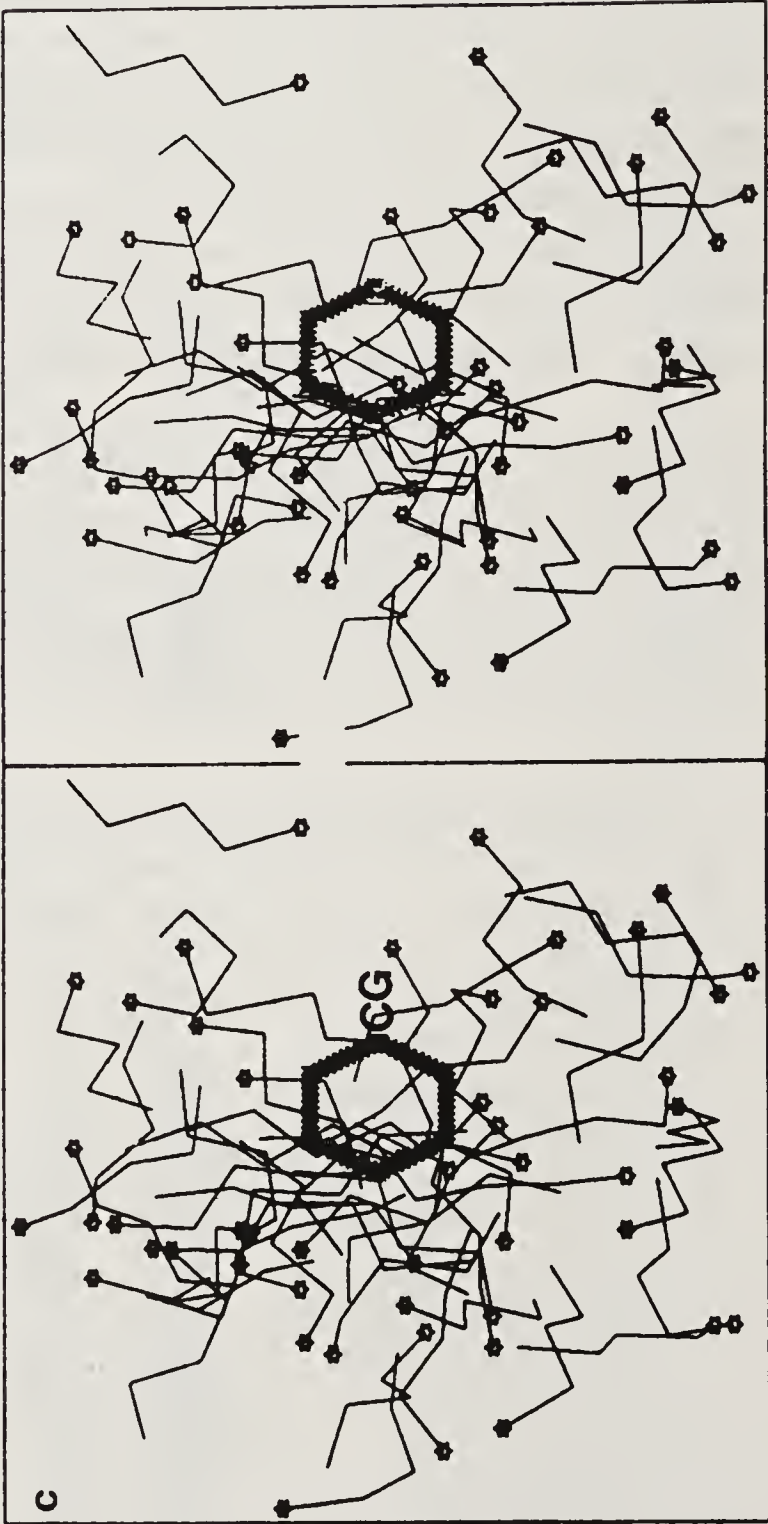
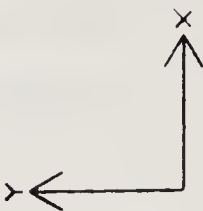
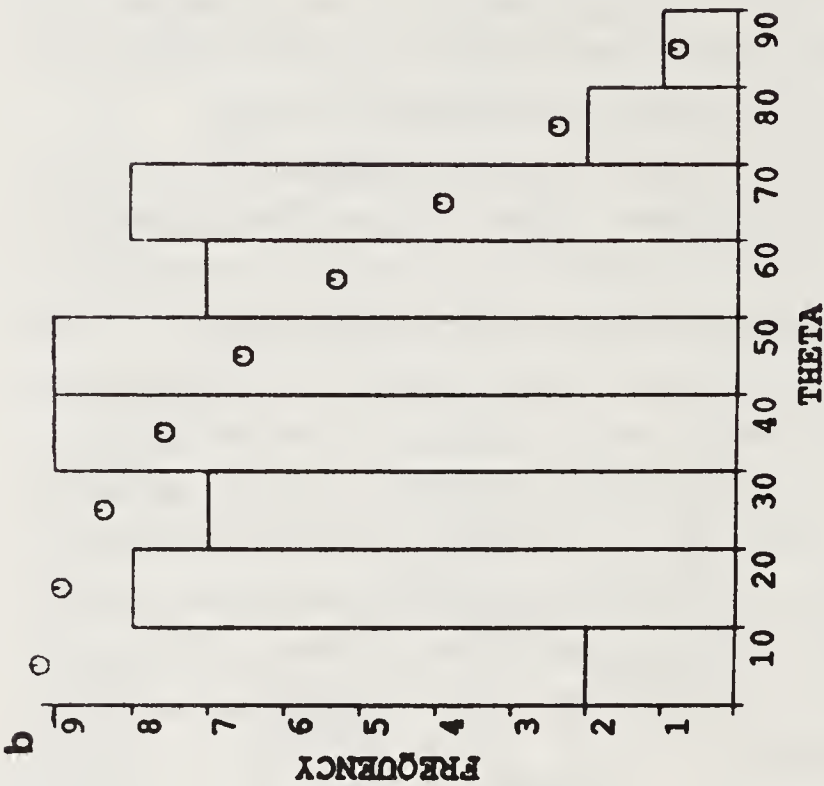
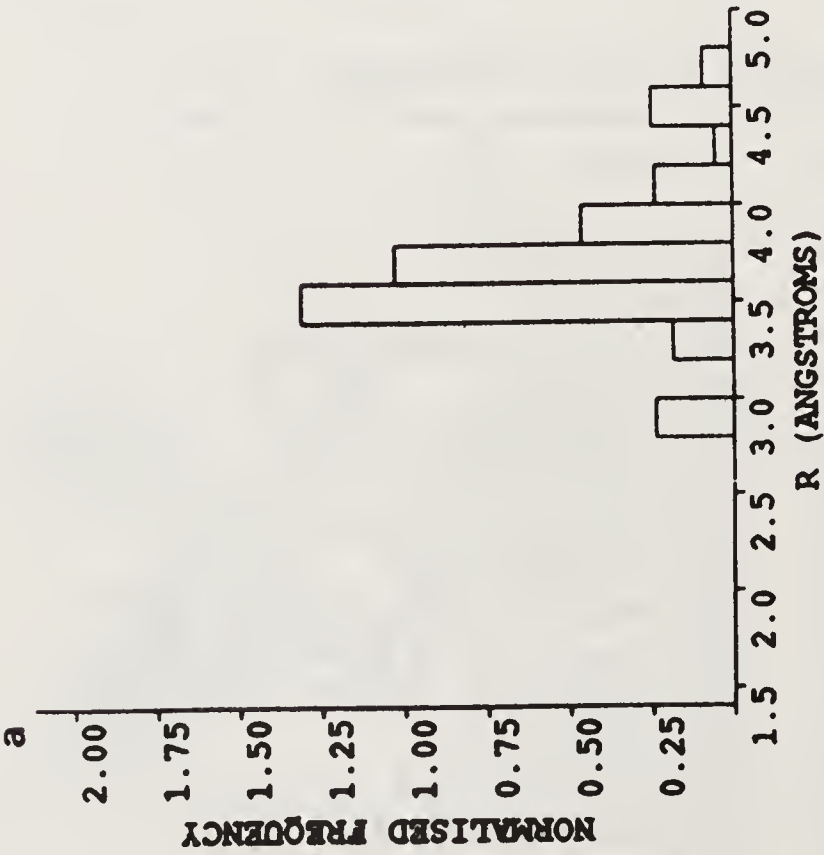
Arg is a polar side chain which is positively charged at physiological pH. Its side chain contains a guanidinium group, with 3 nitrogens. Table 3 shows the frequency of Arg interactions with the different atom types in proteins. Arg–oxygen interactions are common, with all types of oxygen being observed more often than expected. The majority of nitrogens, and the sulphur interactions occur less often, and this suggests they are not favourable. The picture for carbon is more complex. Apolar carbons, aliphatic carbons, polar carbons, backbone C α and backbone carbon are less common than expected. However, aromatic carbons, CH₃ groups, and neutral carbons all are more common than expected. How can we use our database to study side chain–atom interactions? First, we can use it to determine how the packing of carbons, nitrogens, sulphur and oxygen differ and then see how the different types of carbon, oxygen, sulphur and nitrogen compare in their packing.

(I) *Oxygens*: The packing of oxygen from carboxylic acid groups around Arg has been well documented (6, 11, 12). Examination of the other oxygen-containing groups shows similar properties. The carboxyl oxygens, alcohol hydroxyls, and the backbone oxygens all show peaks in the distance profile around 2.8 to 3.2 Å indicative of a hydrogen-bond (figure 7a). In addition to this, the backbone oxygens and the alcohol

Table 3. Observed and expected frequencies of Arg-atom interactions.

Atom type	Total	Nint	Nexp	Examples
Alcohol OH	1911	115	106	S, T(O γ) Y(O η)
Carboxyl oxygens	2140	271	118	D(O δ) E(O ϵ)
Backbone oxygens	10202	571	564	All Backbone oxygens
Carbonyl oxygen	791	59	44	N(O δ) Q(O ϵ)
Carbonyl carbon	1861	208	103	D, N(C γ) E, Q(C δ)
Backbone carbon	10202	439	564	All backbone C – α
Apolar carbon	4246	228	236	(All CH ₂) C, F, I, L, M, P, W, Y
Neutral carbon	2176	141	120	(All CH ₂) H, N, Q, S
Aromatic carbon	1975	138	109	(All ring C) F, H, W, Y
Aromatic CH	4989	446	276	(All ring CH) F, H, W, Y
Carbon CH ₃	5735	350	317	(All CH ₃) A, I, L, M, T, V
Aliphatic CH	2699	120	149	(All CH) I, L, T, V
Polar carbons	4997	271	276	(All CH ₂) D, E, K, R
Backbone C- α	10202	507	564	Backbone carbonyl C
Aromatic NH	961	59	53	H, R, W(N ϵ) H(N δ)
Amide NH ₂	791	56	44	N(N γ) Q(N δ)
Amino NH ₃	1168	47	64	K(N ζ) R(N η)
Backbone NH	10202	242	564	All backbone nitrogens
Sulphurs	453	26	25	C(S γ) M(S δ)

Nint – number of interactions; Nexp – number of expected interactions.



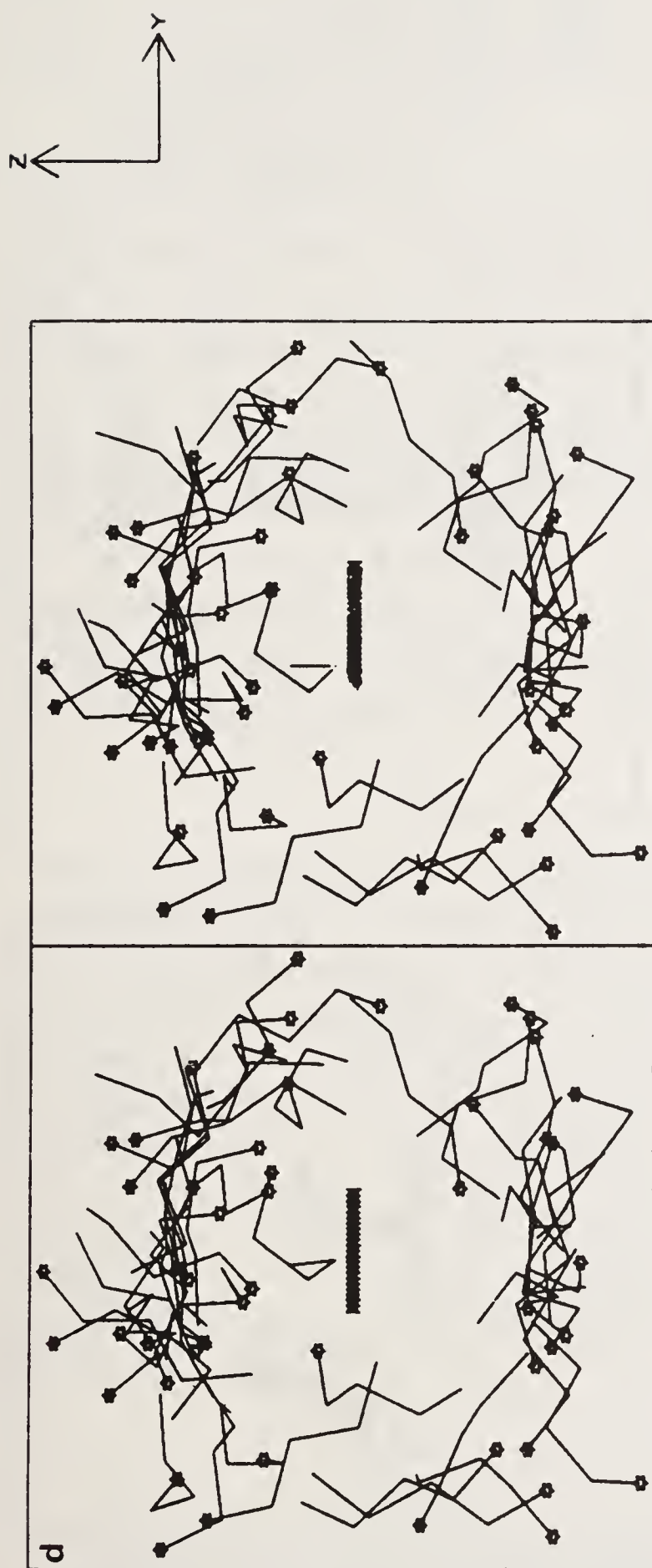


Figure 6. (a) Plot of the closest interatomic distance between the side chains of Phe-Lys pairs in our database. The distance has been normalised to take into account volume considerations. The distance peaks around 3.4 to 3.6 Å as expected from a van der Waals contact. (b) Plot of the angle θ . The angle defines the position of the amino group above the plane of the Phe ring. The random distributions of theta are denoted by a circle. Low values of θ ($\theta = 0^\circ$) indicate amino groups packed against the edge of the aromatic ring, while high values ($\theta = 90^\circ$) indicate amino groups packed above the face of the aromatic ring. (c) Stereo plot of Phe-Lys PAIRS distribution. View along the z-axis towards the face of the reference Phe ring. The amino groups are denoted by stars. It is clear that the amino hydrogens avoid the face of the reference Phe ring. (d) Stereo plot of the Phe-Lys PAIRS distribution. View along the x-axis towards the edge of the Phe ring. The lysines are seen to cluster above the face of the Phe ring, and avoid its edge.

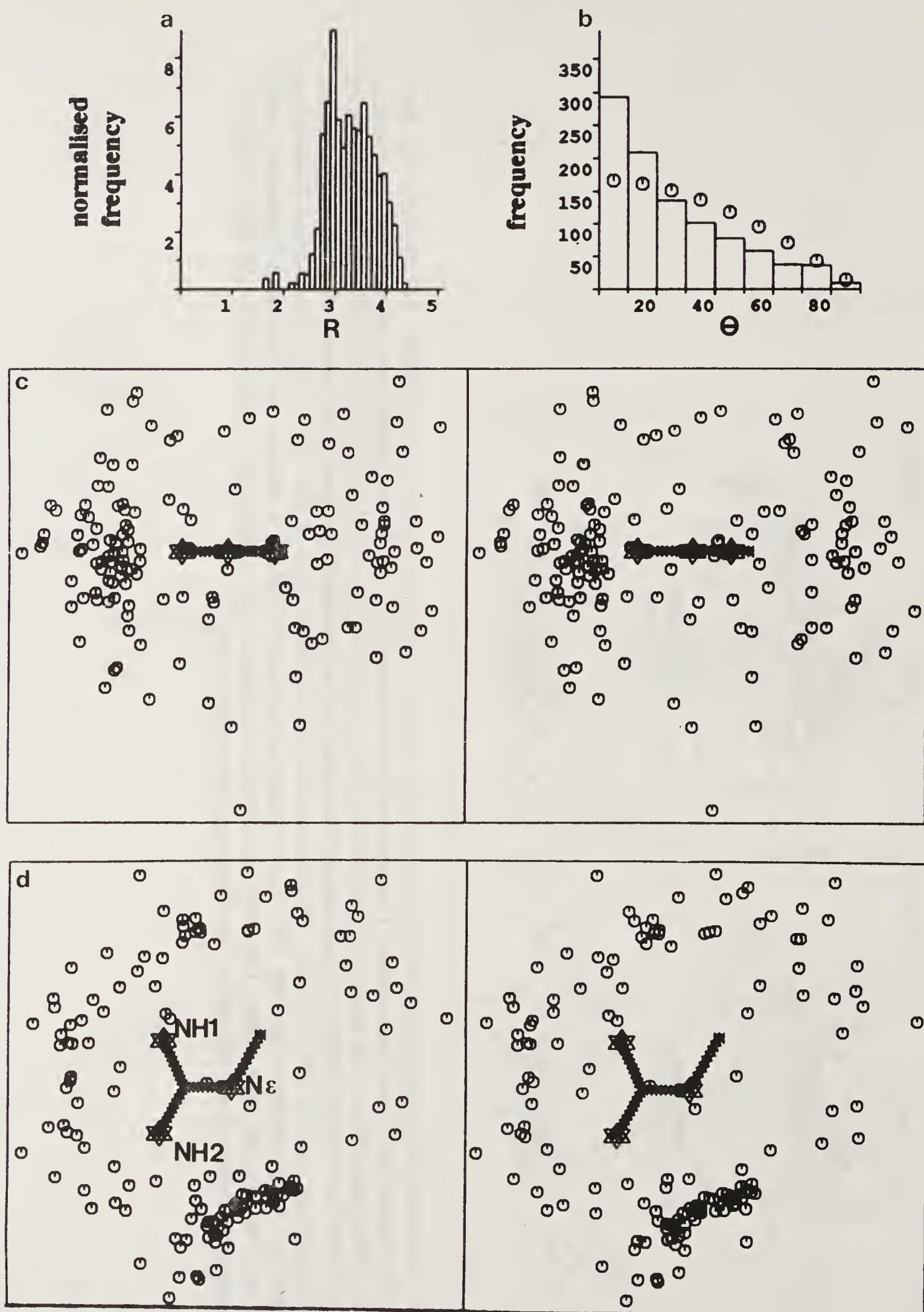


Figure 7. (a) Normalised closest interatomic distance distribution between all oxygen types and Arg side chain atoms. There is a clear peak at 2.8 to 3.0 Å as expected from a hydrogen-bond. (b) Plot of the θ angle of the oxygen position above the plane of the reference Arg. The random distribution is denoted by a circle. There are more observations in the distribution between 0° and 20° , and fewer between 20° and 90° than would be expected by chance. This is a consequence of the geometrical constraint of the hydrogen bond between the oxygen atoms

(Continued on facing page)

hydroxyls show peaks in their distribution around van der Waals distance. These oxygens all share a preference to lie in the plane of the Arg (figure 7b), and to cluster around the nitrogens (figure 7c). There are two possible positions for oxygens to form twin contacts with the side chain nitrogens. These can be with the NH1 and NH2 or the N ϵ and the NH2. Surprisingly, there are fewer oxygens packing against the NH1, NH2 than the N ϵ , NH2, which is a consequence of these interactions being local in sequence. The carbonyl oxygens differ from the other oxygen atoms in that they show no clear preference in R , θ or ϕ .

(II) *Carbons*: The majority of the carbons show a preference to lie in van der Waals contact. The C α and the aliphatic CH being the exceptions with relatively few contacts below 4 Å (figure 8a). The carbons do not adopt the same theta profile (figure 8b and 8c). The apolar carbons, aromatic carbons, aromatic CH, CH₃ and aliphatic CH all show a preference to avoid the plane of the Arg, and stack on its face, while the carbonyl carbon, and backbone carbon show a preference to lie in the plane of the Arg. This is a consequence of their being directly bonded to oxygens which show a preference to lie in the plane. The neutral carbons, C α , and polar carbons are effectively random.

(III) *Nitrogens and sulphur*: Apart from the aromatic NH, there is no preference for nitrogens to pack with the Arg side chain. In addition the θ and ϕ profiles are essentially random suggesting no preferred geometrical arrangement around Arg.

Summary of Atomic Environment of Arg

The above analysis reveals clear patterns of atom packing around Arg. The most interesting result is the clear segregation of most of the carbons and oxygens in space around Arg. The confinement of oxygens to the plane of the Arg is most easily explained by the geometrical constraint of forming a hydrogen bond interaction with the nitrogens of the guanidinium. More difficult to rationalise is the clustering of most of the carbons above and below the plane of the Arg. This could be a consequence of them forming a favourable interaction with the delocalised π system of the Arg, or more simply be a consequence of occlusion of the plane of Arg by the oxygen atoms. In the case of the atomic environment of Arg, all carbons, oxygens, nitrogens and sulphur do not behave the same. The fact that packing preferences are strong at the atom level suggests that the protein does attempt to fold in a manner to optimise the atomic interactions.

(Continued from facing page)

and the Arg side chain nitrogens. (c) Stereo plot of Arg–oxygen PAIRS distribution. The plot is viewing the reference group along the y -axis, towards the x – z plane, looking along the edge of the Arg. The oxygen atoms are denoted as circles, and the Arg side chains nitrogens by stars. There are a few oxygens above the face of the Arg, but many packing with the edge, against the side chain nitrogens. (d) Stereo plot of Arg–oxygen PAIRS distribution. The plot is viewing the reference group along the z -axis, towards the x – y plane, looking towards the face of the Arg. The oxygens appear to preferentially pack around the N ϵ , NH2 side of the Arg, than the NH1, NH2.

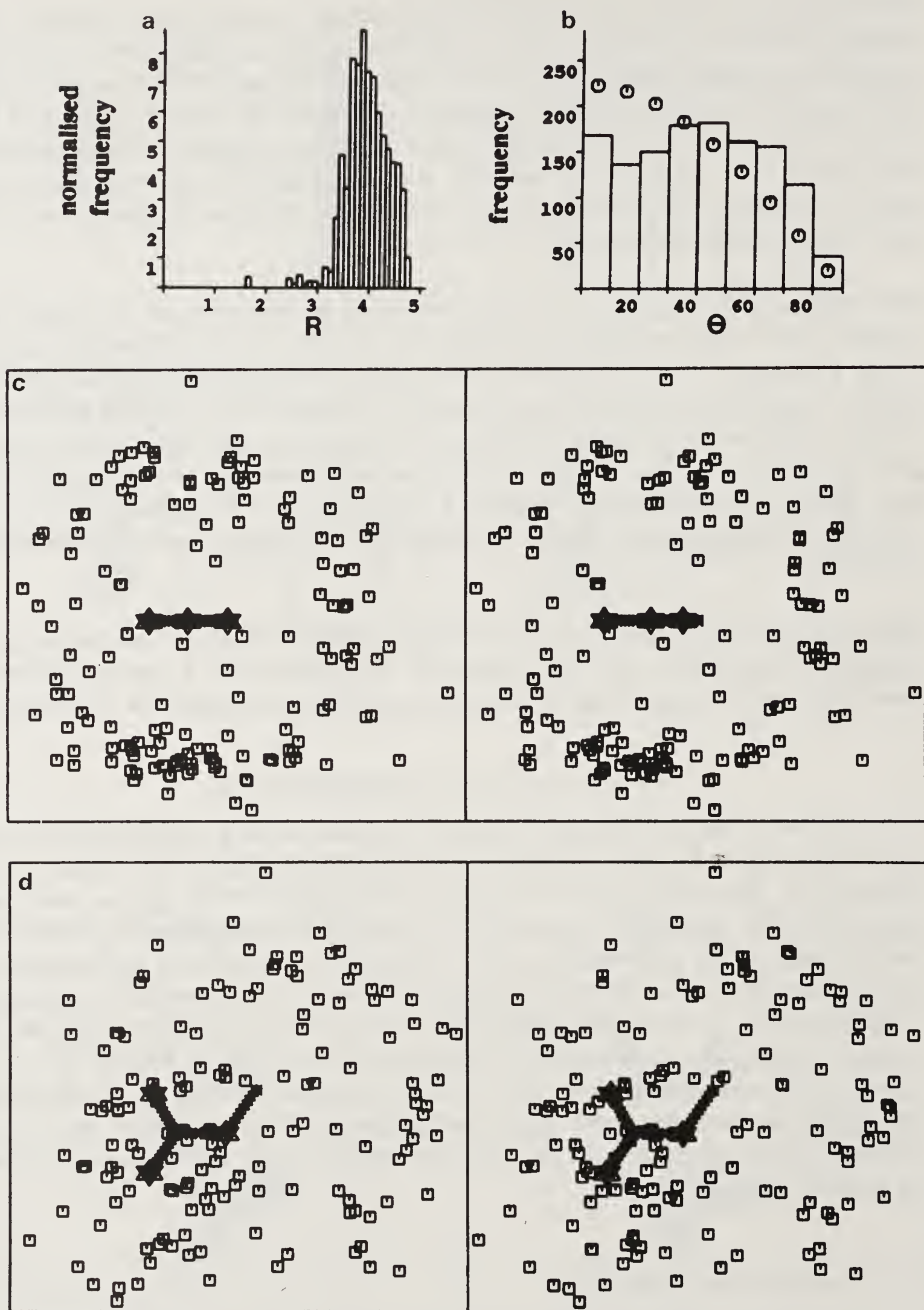


Figure 8. (a) Plot of the normalised closest distance between aromatic carbons, apolar carbons, aliphatic carbons, methyl carbons with the side chain atoms of arginine. (b) Plot of theta angle of position of carbon atoms above plane of Arg. The random distribution is denoted by a circle. Unlike the oxygen distribution around Arg, there are more observations above the plane of the Arg, and fewer below than would be expected by chance. (c) Stereo plot of Arg-carbon PAIRS distribution. Carbons are denoted by squares. Viewing the reference group along

(Continued on facing page)

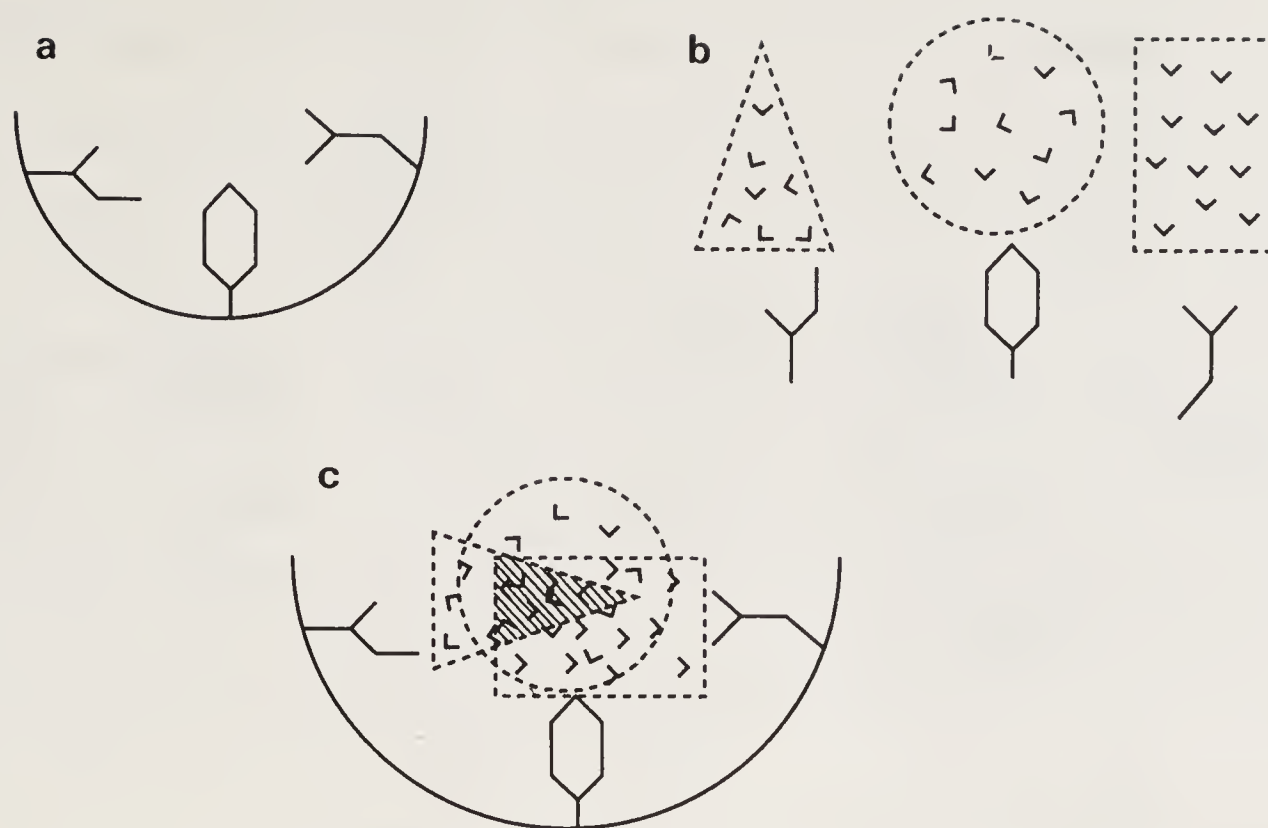


Figure 9. Use of PAIRS distributions in modelling of a carboxyl ligand to a binding site. **(a)** Schematic representation of a receptor site containing an Ile, Phe and an Arg group. **(b)** Schematic representation of 3 PAIRS distributions: Ile-Asp, Phe-Asp and Arg-Asp. **(c)** Schematic representation of 3 PAIRS distributions superposed into the binding site. Involves fitting the reference residue of each of the PAIRS distributions onto the residue at the binding site. The hatched area indicates Asp groups which are common to the 3 PAIRS distributions.

Modelling of Protein Interactions

An obvious application of these PAIRS distributions is to the modelling of protein interactions (13, 14). We have used these data to investigate the feasibility of modelling enzyme-inhibitor interactions. In the majority of cases where the structure of a protein is known, that of the protein-inhibitor complex is unknown. Use of information from existing inhibitor complexes to dock peptide residues into their binding sites is limited by the small number of such complexes available. However, by using the PAIRS distributions we have a wealth of information of protein interactions. One system we have investigated is the binding of the terminal carboxyl group of the dipeptide Gly-L-Tyr to carboxypeptidase A., for which we know the crystal structure. The carboxyl group binds to the S1' pocket, and is involved in contacts with His 69, Asn 44, Arg 145 and Tyr 248. Each of these residues was considered in turn, generating the appropriate pair distribution in the binding site (figure 9). Since there are two side chains containing carboxyl groups, we generated both Asp and Glu PAIRS distributions. For example, for His 69 we generated a His-Asp and His-Glu PAIRS distribution and transformed these onto His 69. For Asn 144, Asn-Asp and Asn-Glu PAIRS distribution was calculated and superposed onto Asn 144, and similarly for

(Continued from facing page)

the y -axis towards the x - z plane. There are few observations of carbons in the plane of the Arg, and many above the face. **(d)** Stereo plot of Arg-carbon PAIRS distribution. The plot is viewing the reference group, along the z -axis towards the x - y plane. There is no clustering of the carbons above the face of the Arg.

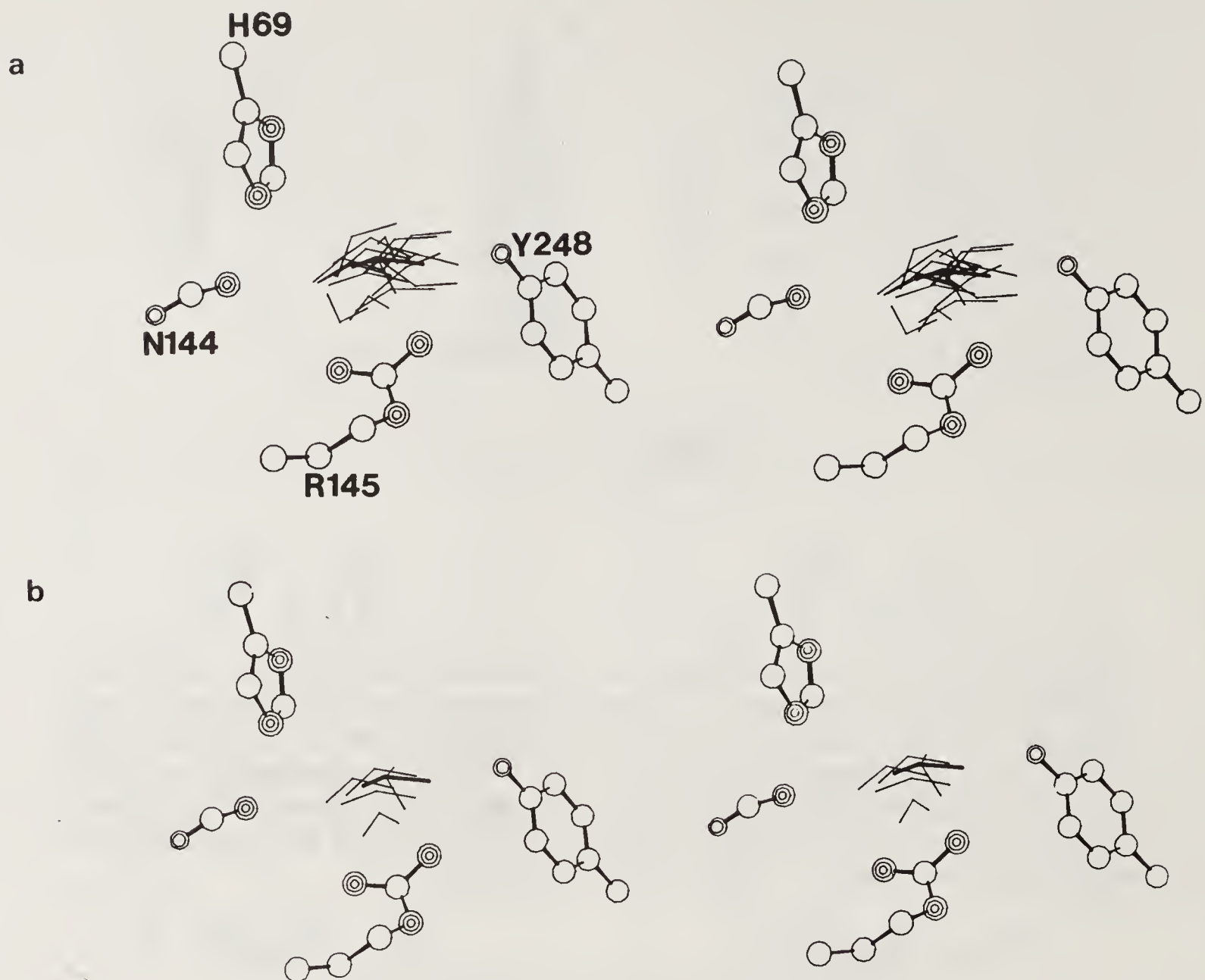


Figure 10. Stereo plots of observed and predicted conformations of the terminal carboxyl group of Gly-L-Tyr. The binding site contains 4 residues His 69, Asn 144, Arg 145, and Tyr 248. The observed inhibitor group is emboldened. **(a)** Plot of the 18 carboxyl conformers identified by overlap of PAIRS distributions. **(b)** Plot of the 5 carboxyl conformers that are able to form reasonable covalent bonds to the C α of Gly-L-Tyr.

Arg 145 and Tyr 248. All carboxyl groups clashing with the host protein side chains were eliminated. In total, for the 4 PAIRS distributions superposed into the binding site, 220 remained after removal of those clashing. Analysis of these 220 possible conformations showed members for which the root mean squared separation from the observed inhibitor group was 1.0 Å. Interestingly, this was only true for residues from the Arg 145 and Tyr 248 PAIRS distributions, and suggests that not all the binding site residues contribute equally to the binding of the ligand. How do we predict these favourable geometries? One method involved identifying overlap between the PAIRS distributions. Since each distribution defines possible packing arrangements of carboxyl groups to each of the binding site residues, where the distributions overlap should define how carboxyl groups should approach the binding site. Using this approach we can reduce the number of possible conformations from 220 to 18 (figure 10a). This can be reduced still further by using information on the backbone location of the inhibitor. Using this approach we can narrow down the number of solutions to 5 (figure 10b). This subset of conformers is of those which are the most similar to the observed inhibitor group.

A Dictionary and Graphics Interface to the Pairs Distributions

Each of the pairwise distributions holds valuable data on side chain packing in proteins. Due to the volume of information, it has been necessary to devise efficient means of examining the data. To facilitate analysis we have developed a dictionary of the data (15). The dictionary consists of 400 pages, each containing three orthogonal stereo plots of each of the PAIRS distributions, and the R , θ , ϕ and P plots including the random expected distributions. This is a useful look-up manual to provide the reader with detailed information on side chain packing. In addition to this static representation of the data, we have interfaced our programs for the display and analysis of PAIRS distributions into a highly sophisticated commercial graphics package. The program allows the user to define the nature of the interacting pairs, and how to display them. An extremely useful advantage of incorporating the data into a graphics program is that they can be used for modelling.

Discussion

It is clear that there is much useful information to be gained from examining side chain packing in proteins for which we already know the three-dimensional structure. At both the side chain and atom level clear packing preferences for these groups are observed in proteins. By studying the side chain–side chain packing geometries it is possible to understand in detail the nature of the interaction between the groups. However the analysis of pairs of side chains is complicated, since it involves multiple atom–atom interactions. To get a simpler picture one needs to study the atomic environment of individual side chains. However, this in turn suffers since one neglects the fact that the atoms are not independent but are connected and also loses information on the orientation of the group. It is therefore important to be aware of both side chain and atom packing geometries. The geometries used here do not completely describe the geometry of the interacting pairs. More rigorous definitions are too complex to analyse. Our approach of a simple geometric definition combined with calculating random distributions of the geometries is sufficiently sensitive to identify preferred packing geometries. It also allows an extensive analysis of side chain geometries. We have gone through the 400 side chain–side chain interactions and calculated the statistical significance of the geometries and are currently interpreting the results. This will then provide a comprehensive source of data on protein interactions.

How may these data be used? The possible applications of these data are considerable because of the role of these interactions in the structure, function and stability of proteins. The general analysis of the side chain data should provide a rule base for understanding the forces of interaction in proteins. Ultimately it will be possible to group the different pair-types on the basis of their geometrical preferences, and from this increase awareness of the energetics of the interaction. The work described here has not attempted to subgroup the interacting pairs according to secondary structure. However, results for aromatic interactions show that there do appear to be preferred packing arrangements in the different secondary structure types (16). It will be important to clarify how the geometries of side chain pairs are dependent on the main chain conformation. This could prove valuable in understanding

how proteins fold. In addition, this data will be used to model protein interactions. This information is currently being applied to exploring and predicting protein–protein, and protein–peptide interactions, and predicting the effects of amino acid replacements in homologous proteins.

Acknowledgements

J.S. is supported by a post-doctoral research fellowship from the SERC Molecular Recognition Initiative. L.C.N. is supported by a SERC CASE award with Pfizer, U.K. We would like to thank Dr. M. Snarey for interest and encouragement during this work. This work is dedicated to the memory of Unkar Singh.

References

1. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T. & Tsumi, M. (1977) *J. Mol. Biol.* **112**, 535–542.
2. Thomas, K. A., Smith, G. M., Thomas, T. B. & Feldman, R. J. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4843–4847.
3. Reid, K. S. C., Lindley, P. F. & Thornton, J. M. (1985) *FEBS Lett.* **190**, 209–213.
4. Singh, J. & Thornton, J. M. (1985) *FEBS Lett.* **191**, 1–6.
5. Burley, S. K. & Petsko, G. A. (1985) *Science* **229**, 23–28.
6. Singh, J., Thornton, J. M., Snarey, M. & Campbell, S. F. (1987) *FEBS Lett.* **224**, 161–171.
7. Singh, J. & Thornton, J. M. (1990) *J. Mol. Biol.* **211**, 595–615.
8. Warne, P. K. & Morgan, R. S. (1978b) *J. Mol. Biol.* **118**, 289–304.
9. Levitt, M. & Perutz, M. F. (1988) *J. Mol. Biol.* **201**, 751–754.
10. Richards, F. M. (1974) *J. Mol. Biol.* **82**, 1–14.
11. Salunke, D. M. & Vijayan, M. (1981) *Int. J. Pept. Protein Res.* **24**, 414–418.
12. Eggleston, D. S. & Hodgson, D. J. (1985) *Int. J. Pept. Protein Res.* **25**, 242–253.
13. Thornton, J. M., Singh, J., Campbell, S. F. & Blundell, T. L. (1988) *Biochem. Soc. Trans.* **16**, 927–930.
14. Singh, J., Saldanha, J. & Thornton, J. M. (1991) *Protein Eng.* **4**, 251–261.
15. Singh, J. & Thornton, J. M. 'An Atlas of Side Chain Interaction in Proteins', in preparation.
16. Singh, J. (1988) PhD Thesis, University of London.

Long and Medium-sized Irregular Regions in Proteins as Combinations of Small Standard Structures

A. V. Efimov

Institute of Protein Research, Academy of Sciences of the USSR, 142292 Pushchino, Moscow Region, USSR

Irregular regions in proteins are here defined as the segments of a polypeptide chain where each residue has, as a rule, a conformation different from that of the neighbors. For example, the loop regions, providing links between elements of regular secondary structure, have irregular but well-defined structures in most globular proteins. There are also many domains and proteins of known three-dimensional structure which practically do not have the regular α -helix or β -sheet secondary structures. Thus, irregular regions form a substantial part of a globular protein molecule and seem to play as important a role in protein folding as α -helical and β -structural regions.

Short loops between α -helices and/or β -strands have been the subject of numerous investigations (1–9). For short loops, there is a restricted number of conformations that may occur, and their conformations depend on specific sequence patterns. Structural determinants of conformations of medium-sized loops in proteins are also considered (10, 11). The main conclusion of this paper is that long and medium-sized irregular regions in proteins can be represented as combinations of small standard structures described in a previous report (4). Since the number of possible combinations is rather large, only some examples of regions and combinations occurring most often in proteins are analyzed.

Classification of Irregular Regions and Description of their Conformations

Classification of irregular regions can be based on their structural features or similar function. There are four types of loop regions if they are classified according to the secondary structures they connect: $\alpha\alpha$, $\beta\alpha$, $\alpha\beta$ and $\beta\beta$. Some loops connect two segments of the secondary structure packed in one layer of a protein sandwich structure as, for example, in β - β -hairpins and in α - α -hairpins. Other loops connect β -strands and/or α -helices localized in the different layers of a protein molecule as in β - β -arcs (8) and α - α -corners (2). Loop regions in β - β -hairpins can be grouped into four classes, depending on the hydrogen-bonding pattern and length (5, 9). Irregular regions, especially short ones, can be grouped into two classes according to their function: turns providing for chain reverse turns by 180° and half-turns to change the polypeptide chain direction by approximately 90° (half-turns are also referred to as

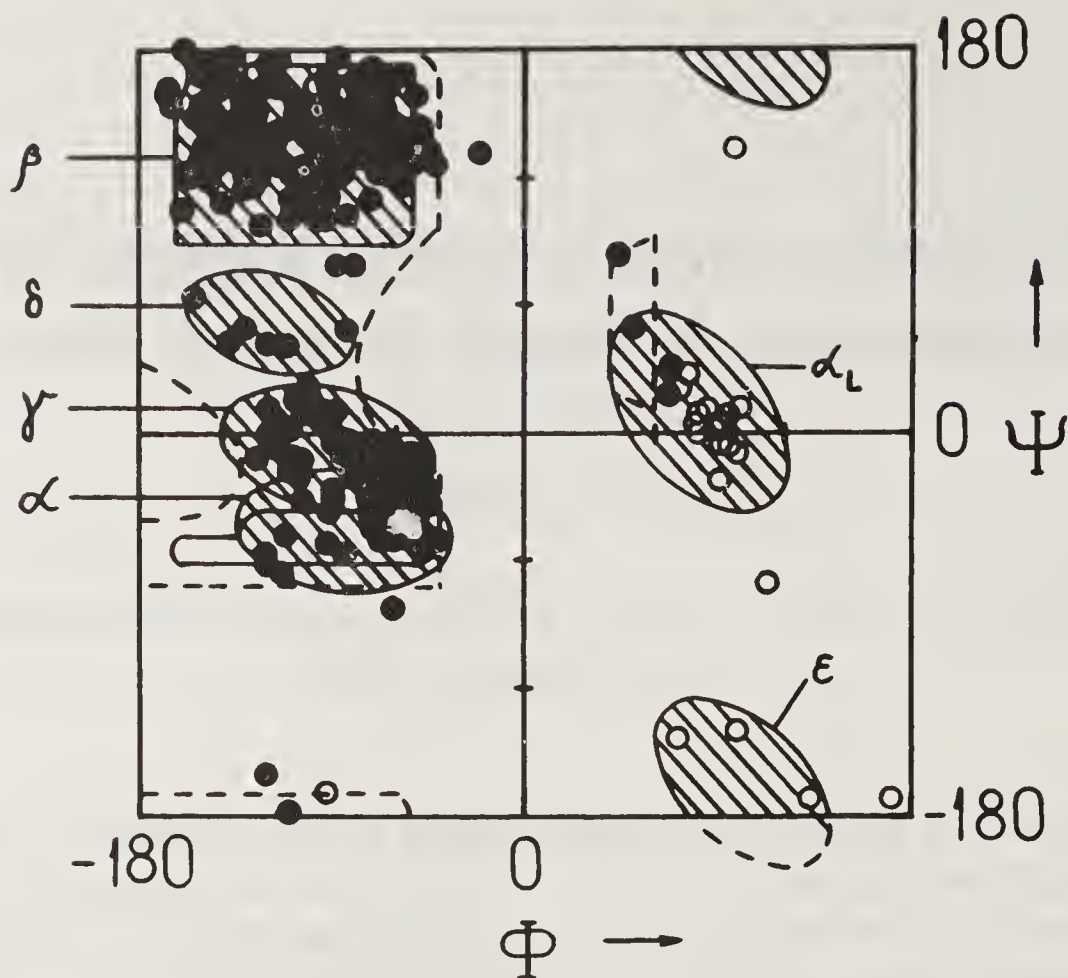


Figure 1. Ramachandran plot for polyalanine (12). Six regions containing the overwhelming majority of experimental points of proteins defined at high resolution are shown by hatching. The ϕ , ψ -angles of non-glycine residues are indicated by solid circles and the glycine residues by open circles. Only the ϕ , ψ -angles for rubredoxin (13), HiPIP (14), plastocyanin (15) and scorpion neurotoxin (16) are plotted.

transitions or cross-overs (4), since they are usually situated at sites where there is a chain cross-over from one layer to another).

Classification based on structural similarity of irregular regions is the most adequate, but comprises many classes. According to this classification, these irregular regions are grouped into one class which almost coincide on superimposition. Irregular regions of the same class have the same length, and the residues occupying equivalent positions have very similar conformations. This classification is of particular value since irregular regions of the same class have similar sequence patterns for the key residues as shown for many loop regions (2–4, 6–9).

To describe the conformation of irregular regions, a shorthand nomenclature (4) is used. This nomenclature has been derived from the observed ϕ , ψ distributions in proteins defined at high resolution (figure 1). As seen, six region, α , α_L , β , γ , δ and ϵ , containing the great majority of the experimental points, can be distinguished. The symbol β_p is also used to designate the polyproline helix conformation. By analogy with the α - and γ -regions, the α_L -region can be separated into α_L - and γ_L -regions but this is not required in the present study.

Small Standard Structures

Small standard structures are shown to consist of three to six residues and can be considered as elements of irregular regions in proteins (4). These are compact structures and most of them are closed into cycles with hydrogen bonds or contacts of

hydrophobic side chains. Such cycling imposes rather strong restrictions on the conformations of the residues.

As mentioned above, small standard structures can be grouped into two classes: (1) $\beta\alpha\gamma\beta$ -, $\beta\alpha_L\alpha_L\beta$ -, $\beta\beta_p\alpha_L\beta$ -, $\beta\alpha\gamma\alpha_L\beta$ - and $\beta\alpha\alpha\gamma\alpha_L\beta$ -turns, providing reverse turns of a chain (some of them are shown in figure 2); (2) $\beta\alpha\beta$ -, $\beta\gamma\beta$ -, $\beta\epsilon\beta$ -, $\beta\alpha\gamma\beta$ -, $\beta\gamma\gamma\beta$ -, $\beta\alpha\alpha\gamma\beta$ -, $\beta\alpha\beta\beta$ - and $\beta\beta\alpha_L\beta$ -half-turns, which provide for only a 90° bend of a polypeptide chain (figure 3). It should be noted that the $\beta\alpha\gamma\beta$ -turn and the $\beta\alpha\gamma\beta$ -half-turn are different structures, although their conformations are described by the same formula (see ref. 4 for details). This demerit of the shorthand nomenclature is explained by the somewhat large sizes of the α , β and γ -regions. On the other hand, a $\beta\alpha\gamma\beta$ -half-turn is very similar to a $\beta\gamma\gamma\beta$ -half-turn as a result of overlapping of the α - and γ -regions.

Turns and half-turns are widespread in proteins (not just in short loops of β - β -hairpins but in long loops and other irregular regions), suggesting that they are stable structures. For many turns and half-turns, residues in the α_L - or ϵ -conformations are required. These α_L - residues must be glycines or hydrophilic residues with flexible side chains while the ϵ -residues must be glycines. The first β -residues of the $\beta\alpha\gamma\beta$ -, $\beta\alpha\gamma\alpha_L\beta$ - and $\beta\alpha\alpha\gamma\alpha_L\beta$ -turns and $\beta\alpha\gamma\beta$ -half-turns should be hydrophilic, Ala or Gly (4).

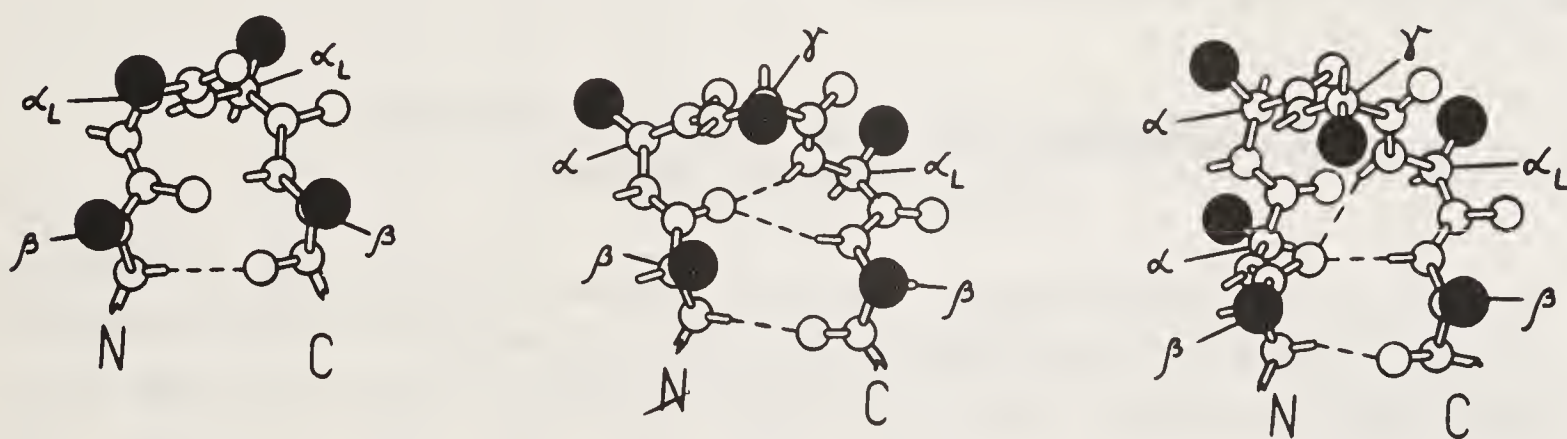


Figure 2. Structures of the $\beta\alpha_L\alpha_L\beta$ -, $\beta\alpha\gamma\alpha_L\beta$ - and $\beta\alpha\alpha\gamma\alpha_L\beta$ -turns shown with ball-and-stick models. Large solid circles are side chains of residues.

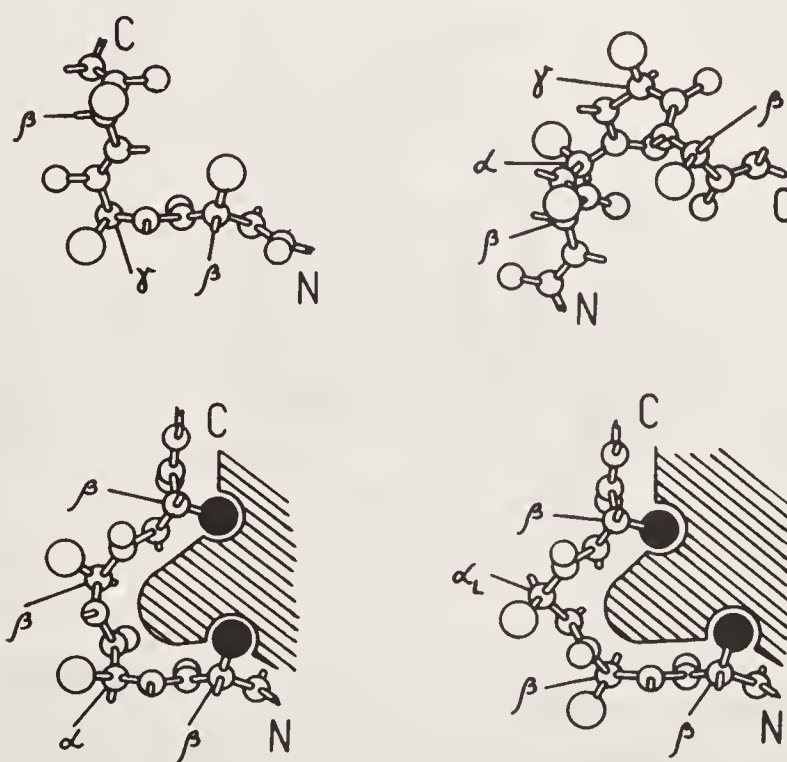


Figure 3. Structures of the $\beta\gamma\beta$ -, $\beta\alpha\gamma\beta$ -, $\beta\alpha\beta\beta$ - and $\beta\beta\alpha_L\beta$ - half-turns shown with ball-and-stick models. Large solid circles are hydrophobic side chains and large open circles are hydrophilics or others. Hatched areas are "hydrophobic cores".

Simple Combinations of Small Standard Structures

As seen in the previous section all the small standard structures have the first and last residues in β -conformations. This property paves the way to combine them: the neighbor standard structures should be connected with one or more β -residues. The connecting β -residues can also be the last residue of one structure and the first residue of the next one simultaneously.

Simple combinations of small standard structures are observed in β - β -hairpin and β - β -arc loops (a β - β -arc is formed by two connected β -strands packed approximately antiparallel and localized in different layers of a protein sandwich structure in contrast to a β - β -hairpin where successive β -strands lie in one layer and are connected by hydrogen bonds). The shortest loop in the β - β -arc has a $\beta\beta_p\beta_p\alpha_L\beta$ -conformation (8). This provides a cross-over of a polypeptide chain from one layer to another and a turn of the chain by $\sim 180^\circ$. But a cross-over and a 180° turn of a polypeptide chain can be provided by two half-turns, for example, by the $\beta\gamma\beta$ - and $\beta\alpha\gamma\beta$ -half-turns, by the $\beta\alpha\beta$ - and $\beta\epsilon\beta$ -half-turns etc. (see figures 4 and 5). Medium-sized loops in β - β -hairpins can also be combinations of two or three small standard structures (see figures 6 and 7). These and other combinations of small standard structures are widespread in loop regions of proteins.

Combinations of Small Standard Structures in Long Irregular Regions

Long loops and other irregular regions in proteins can be formed by three and more turns and/or half-turns. Some examples of such combinations from proteins of known three-dimensional structure are shown in figure 8. It must be noted that in

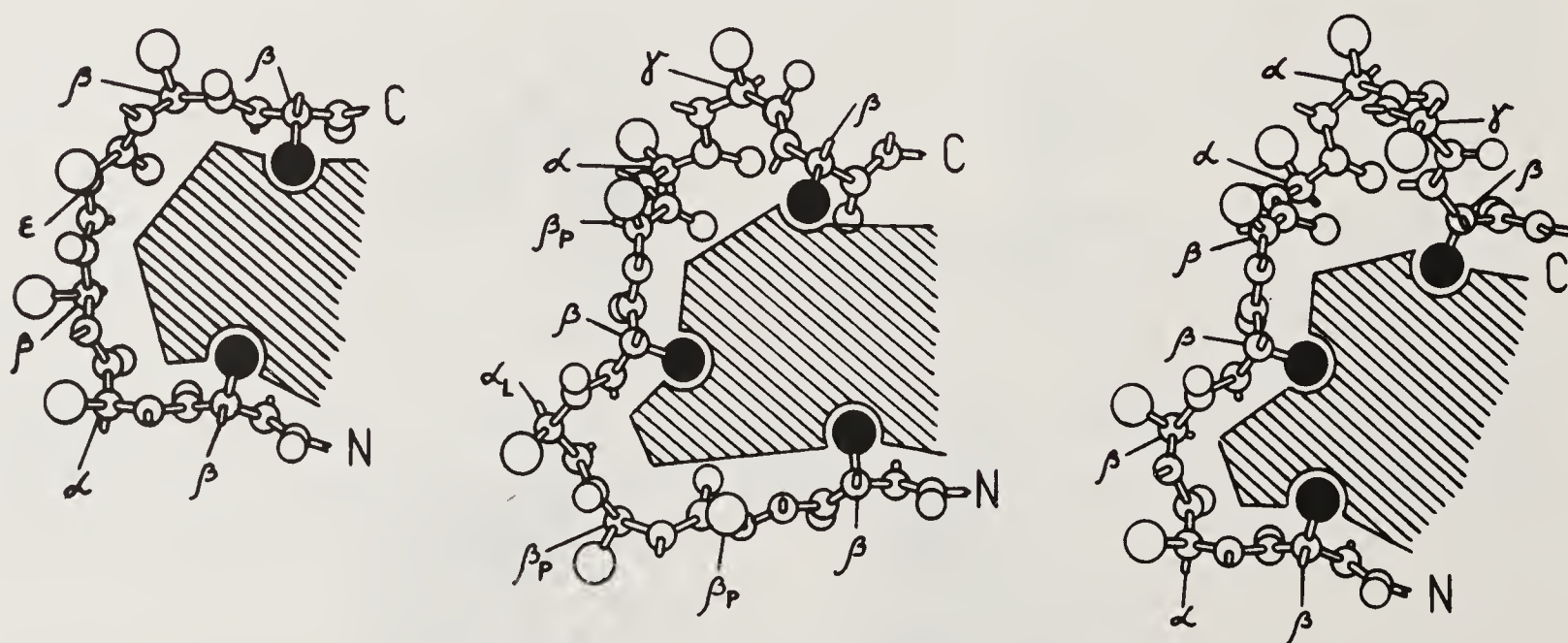


Figure 4. Structures of the $\beta\alpha\beta\epsilon\beta$ -, $\beta\beta_p\beta_p\alpha_L\beta\beta\alpha\gamma\beta$ - and $\beta\alpha\beta\beta\beta\alpha\alpha\gamma\beta$ -combinations of the small standard structures providing cross-overs of the polypeptide chains from one layer to another. Hydrophobic cores are shown as hatched areas. Large solid circles are hydrophobic side chains and open ones are hydrophilics or any other.

figure 8 the conformations of the residues Asp 52, Gly 53, Lys 63 and Glu 64 of ubiquitin are different from those in the X-ray structure of ubiquitin (these residues are labelled with wavy lines). This change has been made since the sequence pattern of the ubiquitin region 50–65 is very similar to that of the plastocyanin region 46–61 and the rubredoxin region 24–37 (see figure 9). The consequence is that all these regions consist of a $\beta\beta_p\beta_p\alpha_L\beta$ -structure followed by an α -helix turn and a $\beta\alpha\gamma\beta$ - or $\beta\gamma\gamma\beta$ -structure.

ILE	PRO	GLY	ALA	THR	CYS					35-40	CRAMBIN
-124	-78	-90	-121	-115	-76	φ					
158	-24	-162	1	105	145	ψ					
β	γ	β	γ	β	β						
VAL	ASP	VAL	GLY	CYS	TRP					37-42A	γ -CRYSTALLIN
-96	-88	-145	69	-104	-130	φ					
111	-48	159	-179	132	158	ψ					
β	α	β	ϵ	β	β						
CYS	SER	THR	ASP	ASN	CYS					57-62	α -COBRATOXIN
-151	-71	-23	-16	-42	-92	φ					
154	-62	90	-73	-33	96	ψ					
β	α	β	α	γ	β						
CYS	GLU	GLY	LEU	PRO	GLU	SER	THR			48-55	SCORPION NEUROTOXIN
-127	-106	75	-76	-69	-59	-82	-82	φ			
133	149	28	153	154	-29	-5	130	ψ			
β	β	α_L	β	β	α	γ	β				
THR	MET	GLU	TYR	ARG	ILE	ASP	ARG			58-65	CI-2
-77	-63	-91	-93	-125	-57			φ			
134	27	128	117	115	-30			ψ			
β	γ	β	β	β	α	γ	β				
LEU	GLN	ALA	GLY	VAL	PRO	SER	ARG	PHE		54-62	BENCE-JONES PROTEIN
-33	-100	-62	66	-87	-66	-92	-58	-85	φ		
125	171	135	4	137	163	-8	-20	133	ψ		
β	β	β_p	α_L	β	β	γ	γ	β			
ILE	SER	SER	LEU	GLN	PRO	GLU	ASP	ILE		75-83	BENCE-JONES PROTEIN
-122	-61	-132	-64	-131	-54	-105	-87	-70	φ		
107	-48	111	143	173	-31	9	-1	118	ψ		
β	α	β	β	β	α	γ	γ	β			

Figure 5. Some examples of cross-overs from crambin (17), γ -crystallin (18), α -cobratoxin (19), scorpion neurotoxin (16), chymotrypsin inhibitor CI-2 (20) and Bence-Jones protein (21). Small standard structures are indicated by open rectangles ().

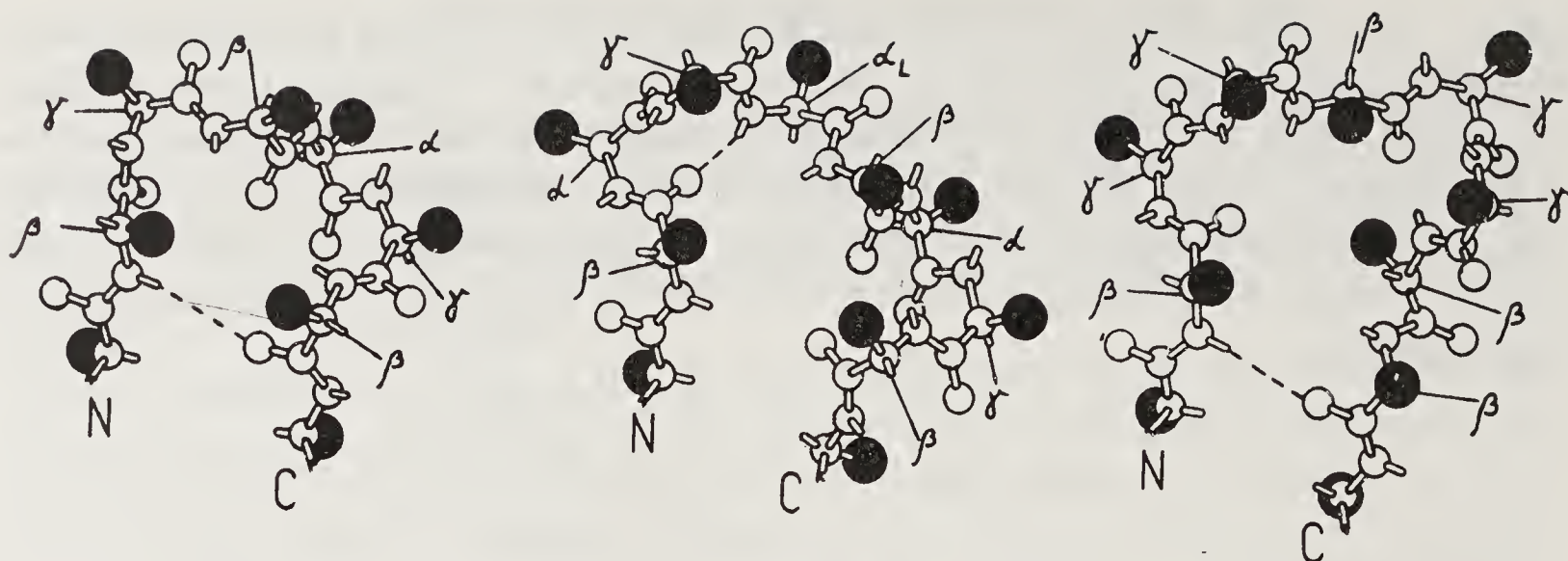


Figure 6. Structure of the $\beta\gamma\beta\gamma\gamma\beta$ -, $\beta\alpha\gamma\alpha_L\beta\gamma\gamma\beta$ - and $\beta\gamma\gamma\beta\gamma\gamma\beta\beta$ -combinations of small standard structures in β - β -hairpin loops shown with ball-and-stick models. Large solid circles are side chains of residues.

LEU -86 130 β	ILE -109 -13 γ	ASN -174 -171 β	GLU -83 2 γ	ASN -131 1 γ	TRP -137 136 β	φ ψ	46-51	CHYMOTRYPSIN		
GLN -136 107 β	THR -81 160 β	PRO -68 132 β_p	GLY 66 2 α_L	LYS -115 180 β	ALA -91 138 β_p	PRO -35 140 β_p	LYS -127 129 β	φ ψ	38-45	BENCE-JONES PROTEIN
MET -99 106 β	ASP -132 32 γ	LYS -70 -19 γ	THR -126 174 β	LYS -79 -2 γ	GLN -123 68 δ/γ	PRO -68 150 β	VAL -130 139 β	φ ψ	11-18	CYTOCHROME C3
SER -45 131 β	SER -73 -14 α/γ	THR -106 -11 γ	CYS 63 29 α_L	SER -46 133 β	THR -86 11 γ	SER -118 -2 γ	THR -107 142 β	φ ψ	217-224	CHYMOTRYPSIN
GLY -148 -168 β	ASP -78 -21 α/γ	GLY -96 173 β	PRO -66 -4 γ	VAL -84 139 β	φ ψ				10-14	SUPEROXIDE DISMUTASE
GLN β	TRP β	ALA α_L	SER β	PRO α	TYR γ	GLY ϵ	ASN β	ALA β	37-45	SCORPION TOXIN II*
SER β	HIS γ	SER γ	THR β	LYS α	VAL β	LYS α	GLU β	PHE β	117-125	DNI*

Figure 7. Some examples of β - β -hairpin loops from chymotrypsin (22), Bence-Jones protein (21), cytochrome C₃ (23), Cu, Zn superoxide dismutase (24), scorpion toxin II (25) and deoxyribonuclease I (DNase I, ref. 26). Proteins labelled with asterisks are those for which conformations of residues are determined taking into account C _{α} -atom stereo plots only.

PHE ASP GLU ASP SER ILE PRO SER GLY VAL ASP ALA SER LYS ILE SER MET SER GLU GLU ASP	41-61	PLA
-92 -74 -67 -88 -138-134-58 -49 89 -83 -84 -58 -62 -73 -117-121-73 -66 -51 -86 -81 φ		
137 114 -23 -3 43 168 141 127 -6 135 104 -31 -29 -27 -7 164 158 152 -37 -12 141 ψ		
β β α γ γ β β β β α α α γ β β α γ β		
PHE ALA GLY LYS GLN LEU GLU ASP GLY ARG THR LEU SER ASP TYR ASN ILE GLN LYS GLU SER	45-65	UBQ*
β α_L α_L β β β β β β α_L β β α α γ α β β α γ β		
VAL VAL ALA GLY GLU PHE ASP GLN GLY SER SER SER GLU LYS ILE	66-80	CHYM
-103-103-114 62 -106-126-101-75 -97 -113-109-100-91 -99 -79 φ		
119 119 155 37 152 -57 113 -15 30 142 -12 16 146 88 160 ψ		
β β β α_L β α β α/γ γ β γ γ β β β		
TRP LEU ILE PRO ASP SER ALA ASP THR THR ALA THR SER THR ASN CYS ALA TYR ASP ARG ILE	191-211	DNI*
β α β β α γ β β β γ β γ β β β α β β β		
VAL LYS ALA GLY ASP THR ARG VAL ILE ALA HIS	73-83	AZN
-144-56 -68 69 -74 -88 -72 -113-73 -161-165 φ		
120 132 133 35 127 -22 -9 118 -41 156 154 ψ		
β β β β α_L β α γ β α β β		

Figure 8. Some long irregular regions from plastocyanin (15), ubiquitin (27), azurin (28), chymotrypsin (22) and DNase I (26). Designations as in figures 5 and 7.

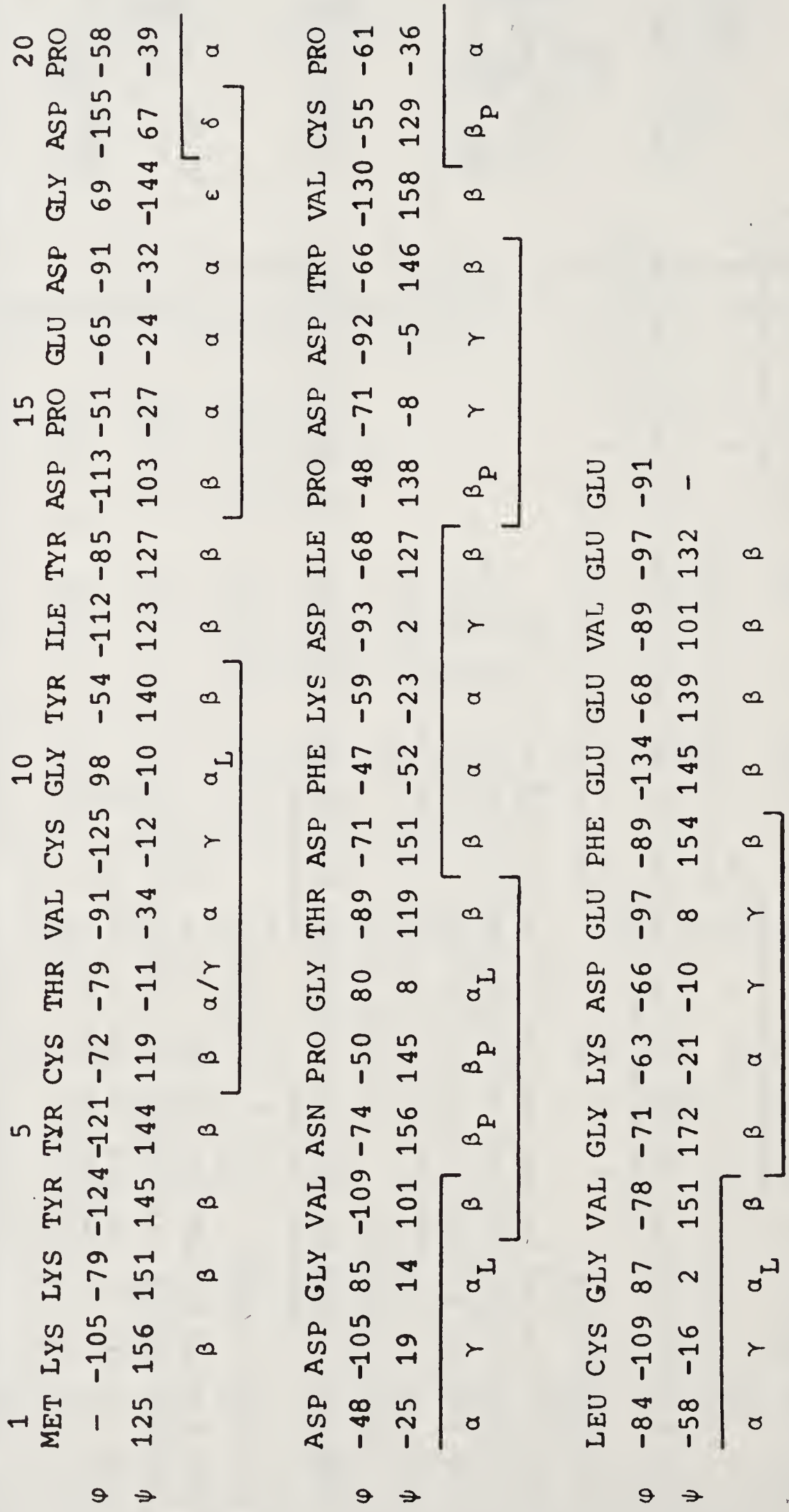


Figure 9. Distribution of small standard structures along the polypeptide chain of rubredoxin (13).

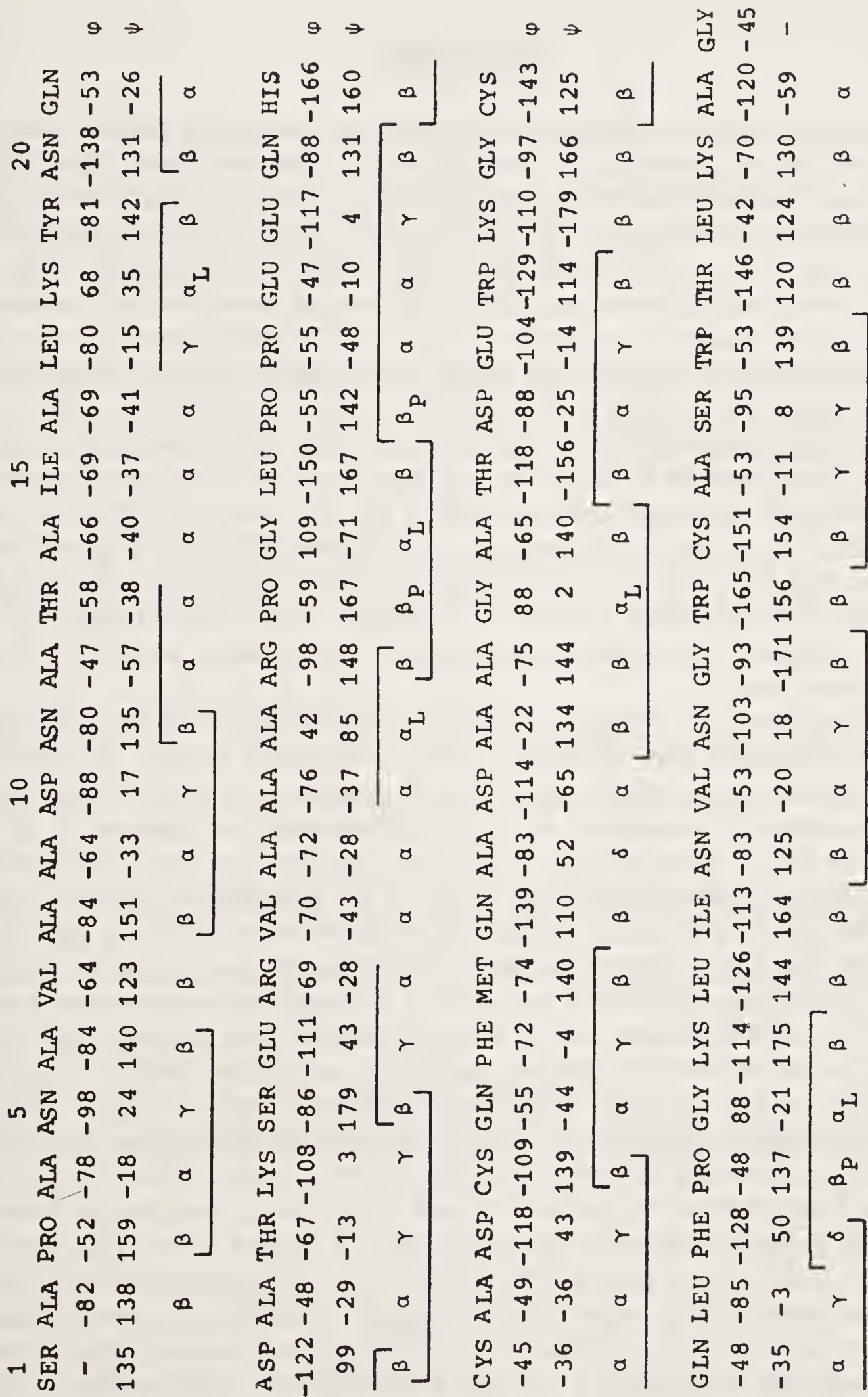


Figure 10. Distribution of small standard structures along the polypeptide chain of HiPIP (14).

Figures 9 and 10 show distributions of small standard structures in rubredoxin (13) and HiPIP (14) molecules where irregular regions form the major part of the structures.

Discussion

The examples shown in figures 5, 7–10 serve as convincing evidence that long irregular regions in protein structures defined at high resolution (1.5–2.0 Å) are combinations of small standard structures, turns and half-turns. Standard structures are connected into combinations with one or several residues in the β -conformations. As a rule, long irregular regions as well as short ones are closed into cycles with hydrogen bonds (as, for example, in long β - β -hairpin loops) or with contacts of hydrophobic side chains or “hydrophobic legs” of hydrophilics (these are usual cases in cross-overs and arc loops). On the whole, the polypeptide chain is closely packed in irregular regions of proteins. In the reverse turns and long loops providing reverse turns of a chain, the polypeptide backbone is tightly packed and side chains are pointed to the solvent. In the half-turns and long loops providing cross-overs of the polypeptide chain from one layer to another (as, for example, in β - β arc loops), hydrophobic side chains or “hydrophobic legs” of hydrophilics are pointed towards each other and form contacts between themselves to form close packing of the polypeptide chain as a whole. If “holes” are formed as in the DNase I region 191–211 (figure 8) or in other metal-binding loops they are occupied by metal ions or other ions and molecules.

An analysis shows that there are no strong steric constraints in irregular regions of proteins defined at high resolution. The overwhelming majority of residues in irregular regions have sterically allowed conformations (see figure 1) and residues having the sterically constrained α_L - or ε -conformations are glycines ($\sim 80\%$) or residues with flexible side chains. All polar groups of a polypeptide chain in irregular regions form intramolecular hydrogen bonds or are accessible to solvent molecules and can form hydrogen or salt bonds with water or other molecules and ions. In other words, there is no dehydration of the free polar groups in irregular regions of proteins. Side chains completely buried in a hydrophobic environment and not accessible to water molecules should be hydrophobic. With this in mind, one can conclude that the order of hydrophobic, hydrophilic and glycine residues in a sequence is one of the main structural determinants of conformations of irregular regions in proteins. For example, the $\beta\alpha\beta\beta$ - and $\beta\beta\alpha_L\beta$ -half-turns shown in figure 3 should have hydrophobic residues in positions 1–4 and Gly in the α_L -position. The $\beta\beta_p\beta_p\alpha_L\beta$ -loop must have hydrophobics in positions 1–5 and Gly in the α_L -position (for details see ref. 8). The $\beta\alpha\beta\varepsilon\beta\beta$ -combination shown in figure 4 should have hydrophobics in positions 1–6 and Gly in the ε -position. The $\beta\beta_p\beta_p\alpha_L\beta\beta\alpha\gamma\beta$ -combination (figure 4) should have hydrophobics in positions 1–5–9 and a hydrophobic or small residue in the first β -position of the $\beta\alpha\gamma\beta$ -half-turn. The $\beta\alpha\beta\beta\beta\alpha\alpha\gamma\beta$ -combination (figure 4) should have hydrophobics in positions 1–4–9 and a hydrophilic or small residue in the first β -position of the $\beta\alpha\alpha\gamma\beta$ -half-turn, etc. The rest of the residues in each combination can be hydrophilic or otherwise.

The structures of irregular regions in proteins depend also on their length, on the secondary structures which they connect and on the relative dispositions of the

connected α -helices and/or β -strands. This means that the structure of an irregular region in a protein is determined not only by its amino acid sequence but also by that of neighboring regions.

References

1. Venkatachalam, C. M. (1968) *Biopolymers* **6**, 1425–1436.
2. Efimov, A. V. (1984) *FEBS Lett.* **166**, 33–38.
3. Sibanda, B. L. & Thornton, J. M. (1985) *Nature (London)* **316**, 170–174.
4. Efimov, A. V. (1986) *Mol. Biol. (USSR)* **20**, 250–260.
5. Milner-White, J. & Poet, R. (1986) *Biochem. J.* **240**, 289–292.
6. Efimov, A. V. (1986) *Mol. Biol. (USSR)* **20**, 329–339.
7. Efimov, A. V. (1986) *Mol. Biol. (USSR)* **20**, 340–345.
8. Efimov, A. V. (1987) *FEBS Lett.* **224**, 372–376.
9. Sibanda, B. L., Blundell, T. L. & Thornton, J. M. (1989) *J. Mol. Biol.* **206**, 759–777.
10. Chothia, C. & Lesk, A. M. (1987) *J. Mol. Biol.* **196**, 901–918.
11. Tramontano, A., Chothia, C. & Lesk, A. M. (1989) *Proteins: Struct. Funct. Genet.* **6**, 382–394.
12. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963) *J. Mol. Biol.* **7**, 95–99.
13. Watenpaugh, K. D., Sieker, L. C. & Jensen, L. M. (1979) *J. Mol. Biol.* **131**, 509–522.
14. Carter, C. W., Jr., Kraut, J., Freer, S. T., Xuong, N. H., Alden, R. A. & Bartsch, R. G. (1974) *J. Biol. Chem.* **249**, 4212–4225.
15. Guss, J. M. & Freeman, H. C. (1983) *J. Mol. Biol.* **169**, 521–563.
16. Almasy, R. J., Fontecilla-Camps, J. C., Suddath, F. L. & Bugg, C. E. (1983) *J. Mol. Biol.* **170**, 497–527.
17. Hendrickson, W. A. & Teeter, M. M. (1981) *Nature (London)* **290**, 107–113.
18. Wistow, G., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P. & Blundell, T. (1983) *J. Mol. Biol.* **170**, 175–202.
19. Walkinshaw, M. D., Saenger, W. & Maelicke, A. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2400–2404.
20. McPhalen, C. A. & James, M. N. G. (1987) *Biochemistry* **26**, 261–269.
21. Epp, O., Lattman, E. E., Schiffer, M., Huber, R. & Palm, W. (1975) *Biochemistry* **14**, 4943–4952.
22. Cohen, G. H., Silverton, E. W. & Davies, D. R. (1981) *J. Mol. Biol.* **148**, 449–479.
23. Higuchi, Y., Kusunoki, M., Matsuura, Y., Yasuoka, N. & Kakudo, M. (1984) *J. Mol. Biol.* **172**, 109–139.
24. Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S. & Richardson, D. C. (1982) *J. Mol. Biol.* **160**, 181–217.
25. Fontecilla-Camps, J. C., Habersetzer-Rochat, C. & Rochat, H. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7443–7447.
26. Oefner, C. & Suck, D. (1986) *J. Mol. Biol.* **192**, 605–632.
27. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987) *J. Mol. Biol.* **194**, 531–544.
28. Norris, G. E., Anderson, B. F. & Baker, E. N. (1983) *J. Mol. Biol.* **165**, 501–521.

Simplified Models for Understanding and Predicting Protein Structure

John M. Troyer¹ and Fred E. Cohen^{1,2*}

Departments of Pharmaceutical Chemistry¹ and Medicine², University of California,
San Francisco, CA 94143-0446, USA

Introduction

Under the appropriate conditions, a protein chain spontaneously folds from denatured, inactive states into an organized functional three-dimensional structure (1). Many experimental techniques, such as NMR spectroscopy and X-ray crystallography, can provide a view of these complex macromolecules at atomic resolution (2,3). Detailed energetic calculations have augmented our understanding of the intrinsic stability and local flexibility of proteins (4). However, an atomic view of large scale conformational changes, including those inherent in protein folding, remains computationally inaccessible. The reasons for this difficulty include: the vast number of conformational states available in principle to a folding chain; the multiple minima which obscure the absolute minimum of potential functions; the small steps required to traverse detailed energy surfaces; and the comparatively time-consuming process of chain assembly (> 1 ms).

The success of simplified models for understanding polypeptide chain conformation began with the seminal work of Ramachandran (5). A hard-sphere model of the conformational preference of a dipeptide accurately revealed the intrinsic local stability of α -helix and β -strand structures. Proteins form locally regular structure and organize this secondary structure into stable folding domains, building complex assemblies from several covalently or non-covalently linked domains. Simplified models which exploit this hierarchical organization can overcome some of the conformational sampling problems inherent in detailed representations of chain energetics, due to a decrease in the number of degrees of freedom which are allowed to vary. A simplified model is conformationally degenerate; many of the very similar conformational substates available to the actual protein become equivalent in the reduced representation. Thus there will be fewer conformationally similar local minima to obscure the global energy minimum. Schematic energy surfaces for an all-atom and a simplified force field are illustrated in figure 1. Conformational searching algorithms using such a simplified force field should be able to traverse the potential surface using longer steps than more complicated force fields.

* For correspondence

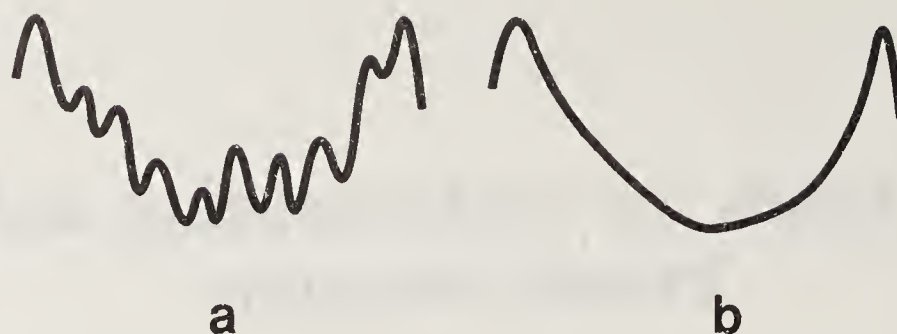


Figure 1. Schematic energy surface for (a) all-atom and (b) simplified macromolecular force field. Many near-equivalent conformational substates become degenerate using a simplified force field.

This chapter will present an overview of current molecular mechanics techniques, and discuss some of their limitations. We will then look at knowledge-based protein prediction strategies, and examine the incorporation of such empirical rules into refinement methodologies for model protein systems. More comprehensive reviews of molecular dynamics simulations and knowledge-based protein modeling are available (4,6,7).

Molecular Mechanics Energy Calculations

Potential energy functions determine relative molecular energies as a function of atomic positions. Such force fields have been developed for small molecules, peptides, and macromolecules (4). Molecular geometry can then be optimized through standard minimization techniques on the multidimensional energy surface. Energy minimization, in all but the most simple cases, will certainly not yield the global energy minimum. The potential energy surface of macromolecules is dominated by many shallow minima which are virtually equivalent in energy due to the relatively 'loose' torsional degrees of freedom. Therefore, energy minimization of a protein is used only with other conformational searching techniques. Various techniques have been developed to search for conformations corresponding to the potential energy minimum (8).

The potential energy functions used in energy calculations usually include independent terms for bonds, bond angles, dihedral angles, 'improper' dihedral angles (to maintain planarity or chirality), and nonbonded interactions such as van der Waals and Coulomb terms. Hydrogen bonding is treated electrostatically, although some force fields include explicit hydrogen bonding terms to optimize the donor-acceptor distance. More complicated force fields for the optimization of small molecules, such as MM2, include combined bond bending-stretching terms in their potential functions (9).

The force fields used for macromolecules contain many approximations (9). One could in theory determine structural energies from quantum mechanical calculations, but this is computationally prohibited for all but the smallest molecules. A general assumption held in the development of macromolecular force fields is that parameters are transferable between systems. Parameters derived from spectroscopic and quantum mechanical studies on small molecules are applied to equivalent fragments of large molecules. The energy functions are chosen so that they and their derivatives are simple enough to calculate easily and quickly, but are of an appropriate form such that the

correct energies result over the range of configurations of interest. For instance, some accuracy is lost through the neglect of many-body nonbonded interactions, but pairwise terms are parameterized to account for this lack. The neglect of bend-stretch interactions in many of the current macromolecular force fields is likewise taken to have only small effects over the gross conformational space traversed.

A force field can be coupled with molecular dynamics (MD) algorithms to sample local conformation space around the energy minimum of a macromolecule (4). The MD algorithm involves the numerical integration of Newton's equations of motion, where $a_i = F_i/m_i$, and the force F_i is the negative gradient of the potential surface. New atomic positions and velocities are calculated repeatedly with a small time step on the order of 10^{-15} seconds. The resulting set of time dependent coordinates is known as a trajectory.

The most time-consuming task in each iteration is the evaluation of nonbonded interactions. In theory, all pairwise interactions must be computed, and the number of interactions varies as the square of the number of atoms in the system. Use of a cutoff radius, beyond which interactions are assumed to be negligible, is expedient. However, the energy of the system is not conserved due to this abrupt truncation. 'Hot spots' can develop, especially in the solvent (D. Spellmeyer, P. Kollman, personal communication). Since simulations are often carried out at a constant temperature by coupling to a temperature bath, uneven heating is partially mitigated. Switching functions that scale these interactions gracefully to zero can be used to avoid these problems (10,11).

The time step used in MD calculations must be approximately one order of magnitude smaller than the highest frequency motion of the system. Usually, these are bond-stretch vibrations, ($\approx 10^{14} \text{ s}^{-1}$), which limit MD time steps to approximately one femtosecond. In order to sample more of conformational space, bond lengths can be constrained to be constant without appreciably changing the dynamics of the system. This speeds up calculations by a factor of three (12). Bond angles are more strongly coupled with the dihedral angle degrees of freedom, and may not be constrained without decreasing atomic fluctuations by a factor of two and strongly inhibiting dihedral transitions (12).

Molecular dynamics permit a view of the structure and stability of macromolecules that is not obvious from the examination of static macromolecular structure. Qualitative insights into protein structure can also be gained through molecular dynamics. Wendolowski and coworkers carried out MD simulations to investigate electron transfer in the cytochrome c–cytochrome b_5 intermolecular complex (13). During the course of the calculations, fluctuations in the position of side chains and the two heme groups were consistent with the facilitation of electron transfer. Koehler *et al.* carried out studies of hydrogen bonds in crystalline and aqueous systems that were consistent with experimental results (14). They suggest that bifurcated hydrogen bonds may be intermediates in changing hydrogen bond configurations. Makinen and coworkers correlated the simulated mobility of regions of the carboxypeptidase A active site with their catalytic function (15). Molecular dynamics can be coupled with constraints from experiment, such as interatomic distances derived from NMR NOE data (16–18) or position restraints from the crystallographic electron density (19). This restricts the region of conformational space to be searched, and allows a structure consistent with experimental results to be defined. MD has also proved to be a useful tool in designing and interpreting protein engineering studies (20, 21).

Molecular dynamics, while incorporating flexibility into the static picture normally considered in macromolecular structure, does not intrinsically incorporate entropic factors into its energy calculations. Entropy can be taken into account using the ergodic assumption that the trajectory has traversed all of phase space and is equivalent to a thermodynamic ensemble of states. Using thermodynamic perturbation or integration methods, differences in free energies can be calculated. For all but the simplest perturbations, the ensemble of states is unlikely to be complete or even representative of a true ensemble. Despite this fact, free energy calculations can agree well with experiment (22). A recent review of the results of free energy perturbation calculations has been compiled by Beveridge and DiCapua (23).

The longest available protein trajectories span several hundred picoseconds. Even though increasing computational power is allowing ever longer trajectories to be calculated, the efficiency with which molecular dynamics samples conformational space must be considered. Since MD produces a Boltzmann-weighted trajectory in phase space, high-energy transitions between two stable conformers will occur quite infrequently. Brooks indicates that to adequately sample conformational space for even an eight-residue peptide, multiple high temperature (600 K) long (1 ns) trajectories may be necessary (24). Larger systems remain computationally inaccessible at present. Much of the available phase space for a protein will not be sampled by a single nanosecond trajectory. It appears that, even for small, highly constrained molecules, MD samples conformation space inefficiently when compared to other searching methods (25).

It is possible that more effective sampling protocols, such as simulated annealing, may be developed for macromolecular conformational searching. Although simulated annealing has been used in conjunction with molecular dynamics and Monte Carlo methods, it is usually carried out in systems with other constraints (18, 26). These experimental constraints prevent the protein from moving too far away from a starting conformation at high temperatures. Simulated annealing has been successful in finding low energy conformations for amino acid dipeptides (27).

Monte Carlo techniques have also been used to sample molecular conformations. Using the Metropolis algorithm, conformations are selected using a probabilistic method and subject to a Boltzmann distribution of energies. With the appropriate choice of variables, this provides a useful strategy for exploring macromolecular structure. Go and coworkers have used these techniques, making several additional assumptions, to simulate motions in proteins (28, 29). Because of the large dimensionality of protein simulation, the only degrees of freedom used in the system are dihedral angles; bond lengths and bond angles are kept constant at their ideal values. Consequently, atomic fluctuations in a Monte Carlo simulation of BPTI were half as much as those from an MD simulation of the same molecule (28).

The use of cartesian coordinates or internal angles as variables in Monte Carlo algorithms leads to very inefficient conformational sampling (30). Most changes in these variables result in very high energy conformers, due to abnormal bond lengths and high van der Waals overlap. Instead, collective variables representing the motion of many atoms are used. These variables correspond to the calculated normal modes of the system (31). In order to calculate normal modes, the energy surface is assumed to be quasi-harmonic (32). Since this approximation is valid only near energy minima, the normal modes must be recalculated at intervals throughout the calculation. Using these assumptions, the Monte Carlo procedure is claimed to be 50 times more efficient

than MD. Although proteins simulated with this technique do explore different conformational substates (29), the effects of the approximations made have not been fully explored.

Frequently, MD studies are performed *in vacuo* to reduce the computational burden. Well-defined waters from X-ray studies are included occasionally. Of course, macromolecules in physiological systems are solvated. To correctly model these systems, it is necessary to include the free energy of solvation in the calculation. This cannot be accomplished with static modeling of water due to the omission of entropic effects. One effect of including water in a simulation is to increase the flexibility of solvent-accessible sidechains (33,34). Such solvent effects may be of importance in active site structure and ligand binding (34–36). The inclusion of solvent in simulations has led to a better correlation with experimental results. Using explicit solvent, free energy calculations of the *cis*–*trans* energy difference of the peptide bond (37), the phi–psi map of the alanine dipeptide (38), and the energy barrier for tyrosine rotational isomerization (39) were all in agreement with experimental measurements.

In a full solvent representation, the molecule of interest is placed within a large box of water molecules. Periodic boundary conditions are employed to eliminate boundary effects. A dilute solution is achieved by making the box dimensions greater than the sum of the protein diameter and the nonbonded cutoff distance. Thus, no protein intermolecular interaction occurs. The size of this system is much larger than the protein *in vacuo*, with a resultant decrease in calculation speed. It is also possible to place only a shell of solvent molecules around the protein with no boundary conditions. Although edge effects at the solvent–‘vacuum’ interface can perturb the system, surface residues can interact with solvent molecules without the computational cost of a full box of waters. A third alternative is to include only a cap of waters surrounding a region of interest, such as an enzyme active site. To avoid edge effects, a stochastic boundary region can be included in the calculation. Random forces are applied to atoms in the region to simulate the effects of the surrounding atoms of the protein and bulk solvent (34).

Electrostatic interactions are among the most important factors in determining protein structure and function. Unfortunately, electrostatics are not treated rigorously in current force fields. This is because electrostatic interactions are complicated functions of many atoms that act over long ranges compared to other terms in the molecular mechanics force field (40, 41). Explicit inclusion of solvent with the appropriate interaction parameters may provide an adequate treatment of electrostatic behavior in proteins. The convergence times for such a system will increase greatly, though, due to the long diffusional relaxation times of the solvent. The treatment of induced polarization effects would increase accuracy, but at the cost of increasing computational demands (42, 43). To simulate the charge screening effect of solvent without its explicit inclusion in protein simulations, an effective dielectric is often used (4). This is often in the form of a distance-dependent function. However, the interaction energy of point charges will depend on the position of the charges in the protein and the shape of the protein as well as their separation (44). The use of an effective dielectric constant or function has been called ‘ill-defined’ and ‘confusing’ (44, 45). Despite this, there have been many estimates of an appropriate dielectric constant to use in protein molecular mechanics calculations (46–48). Wendolowski and Matthew have shown that during the course of MD simulation, pK values of individual sidechains can vary by up to 2 pK units (49), indicating that a dynamic treatment of electrostatics is

necessary, and may be especially important in the consideration of catalytic events. The incorrect handling of electrostatics remains a major source of error in current macromolecular simulation.

Knowledge-Based Modeling

Explicit molecular mechanics solutions to the protein folding problem remain many orders of magnitude beyond current computational limits. Even local conformational subsearches around a folded, equilibrium structure are computationally expensive. Experimental results suggest that the 58-residue bovine pancreatic trypsin inhibitor (BPTI) folds in under 10 seconds (50). A complete simulation of the folding process would take 10^8 years of CPU time under the optimistic assumption of 10 ps of simulation for each CPU hour of computation. We can expedite this process by searching for principles that are important in the folding and structural stability of globular proteins and then implementing these principles in structural calculations. There are currently over 300 protein structures in the Brookhaven Protein Data Bank. Approximately 100 of these are high-resolution and mutually non-homologous structures. These experimentally observed 'solutions' to the protein folding problem contain information on the thermodynamics of chain folding.

Among the forces that determine globular protein structure, the partitioning of a relatively hydrophobic core of amino acids from a more polar exterior is thought to be important (51, 52). Surveys of known protein structure demonstrate that globular proteins conform with this partitioning. Many semi-empirical approaches for the prediction of protein structure rely on this paradigm. In an attempt to quantitatively exploit this information, several scales of amino acid hydrophobicity have been developed (53). These are based on experimental data for isolated amino acid side chains, or more typically the distribution of residues in proteins of known structure. Chothia showed that the free energy required to transfer an amino acid side chain from octanol to water was directly proportional to its accessible surface area (54). Eisenberg and McLachlan calculated free energies of solvation for individual atom types as a function of accessible surface area (55). Relative free energies of transfer calculated as the sum of these atomic free energies correlate well with observed free energies.

In principle, the atomic free energy of solvation could be combined with an analytic definition of accessible surface area and incorporated into a molecular mechanics force field. This term would approximate the impact of multiple solvent conformations without the explicit inclusion of solvent molecules. The calculation of accessible surface area using numerical integration can be a computationally intensive task (56, 57). Wodak and Janin developed a simple analytic approximation of residue solvent accessible surface areas. This probabilistic formulation was remarkably accurate when summed over a complete protein and was accurate to within 20% for individual residues (58). However, the detailed behaviour of this function and its derivatives cast doubt on its utility in semi-empirical polymer simulations. Analytic representations of the solvent accessible surface area using techniques from differential geometry have been developed (59, 60). In addition, the derivative of the accessible surface area with respect to atomic coordinates can be calculated from these functions. These formulations remain computationally taxing. Efforts are underway to implement these methods in existing packages (D. Eisenberg, personal communication; P. Kollman,

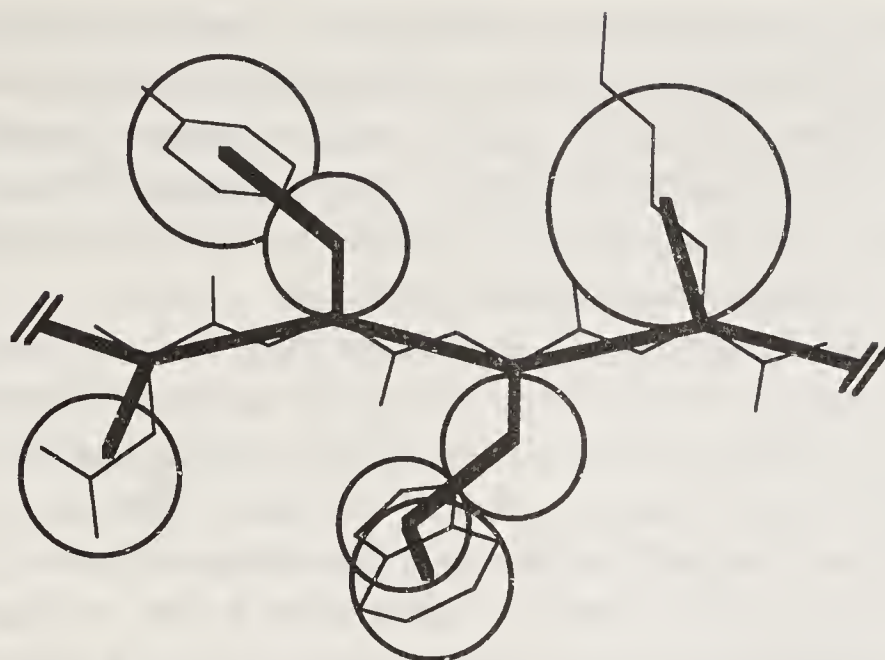


Figure 2. The simplified protein representation used in the QPACK model evaluation procedure for the sequence Leu–Tyr–Trp–Lys. The all-atom polypeptide structure is depicted using thin lines. This simplified model is shown with thick lines. Aliphatic residues are represented by pseudoatoms at their side chain centers of mass. Aromatic residues have additional pseudoatoms placed at their ring centroids.

personal communication), but it is not known if this approximation will be effective in reproducing the effects of solvent on protein dynamics in detailed molecular systems.

The interiors of proteins are well-packed (61). Depending on the size of the van der Waals atomic radii chosen, hard-sphere models of crystallographically-determined protein structures reveal few cavities (56, 57, 62). Some of these cavities contain water molecules in the crystal. We have developed a method to rapidly evaluate packing density in model structures. A simplified residue approach is used, with residues represented by one to three pseudoatom spheres. The algorithm is implemented as the program QPACK written in the C programming language (63). Figure 2 illustrates this simplified protein model. Each type of amino acid is given an “ideal radius” based on its average packing environment taken from a set of 72 high-resolution structures from the Brookhaven Protein Data Bank. This method emphasizes the homogeneity of packing in proteins and is sensitive to changes in orientation of a single side chain. It can differentiate between the correct crystal structure of flavodoxin; a structure built with the flavodoxin amino acid sequence and backbone, but with side chain placed in their most common conformations; a structure with a jumbled flavodoxin amino acid sequence placed on the flavodoxin backbone; and several model flavodoxin structures which optimize secondary structure packing but neglect the connecting loops.

It has long been recognized that proteins with similar amino acid sequences have similar three-dimensional protein structures. Differences between homologous structures usually occur in loops on the protein surface (64, 65). The framework or core residues are conserved owing to constraints on the composition of interior sidechains and other structural restraints (66, 67). When a sequence is homologous to a protein of known structure, a knowledge-based approach to modeling is a very powerful tool for the prediction of the new protein’s structure (7). Unfortunately, recognizable sequence homology is lost more rapidly during evolution than the underlying structural similarity.

Protein segments assume only a limited number of conformations in folded proteins (68, 69). This limited dictionary of templates may be due to the inherent steric constraints of the backbone and other packing and electrostatic interactions in protein

folding. This was most graphically demonstrated by Jones and Thirup. They showed that the majority of the polypeptide backbone of a new structure can be built up from component pieces of other structures, whether or not they are related (68). Unger *et al.* (69) showed, in fact, that 76% of the main chain structure of an 82-protein dataset can be constructed using only 100 unique 6-residue templates. Amino acid side chains are also found in a restricted set of conformers (70, 71).

Because of the limited dihedral space available to protein backbones and side chains, it is possible to construct *de novo* limited polypeptide segments using either a restricted rotational isomeric model or an energy-biased conformational search. Semi-automated algorithms have been developed to carry this out by Scheraga and coworkers (72, 73), Moult and James (74), Levinthal and coworkers (75), and Bruccoleri and Karplus (76). Computation time for more than a five- or ten-residue segment becomes prohibitive. Both database template matching and build-up procedures are well-suited to the construction of short, non-homologous loop segments, and succeed quite well at arriving at a small number of reasonable conformations.

Using the concept of "canonical structures", Chothia and coworkers have succeeded in modeling the hypervariable regions of heterologous immunoglobulins (77). The conformation of the hypervariable loops seems to be determined to a large degree by the interactions of certain key residues at specific sites. When model loops are compared to crystallographic structures subsequently determined, the rms deviation has been less than 1.5 Å. It is important to note that the structural alignments developed in this work do not always agree with standard alignment results. Presumably this relates to the variations in evolutionary pressure along the sequence.

Enumeration of the limited number of conformations available to proteins has been used to advantage in generating loop geometries. While there are many global constraints on possible loop structures, those that satisfy local constraints provide a useful starting point. When these local solutions are combined, the number of possibilities expands combinatorially. Explicitly considering each possible combination is prohibitive. In general, the set of all combinations must be pruned either by a set of filters to reject inconsistent or unlikely results or by an energy function which eliminates high energy structures or substructures. This truncation gives reasonable solutions to the problem, but is somewhat arbitrary. While the prospect of building a novel protein structure from fragments remains tantalizing, in practice, tertiary restraints have been necessary to achieve a reasonable structure. These restraints have come from the framework structures of homologous proteins, long range interatomic distances as measured by NMR spectroscopy, or the molecular envelope of a crystallographically-determined electron density.

A variation on the fragment approach is to explore the chain for potential periodic structures and subsequently investigate the packing of units of secondary structure. Since regions of secondary structure have conformations in folded proteins close to their isolated minimum-energy structures, the packing of helices and β -sheets can be understood using simple geometrical models. These interactions are largely hydrophobic in character and are conformationally restricted by packing at the interface between the units of secondary structure (78). The geometric details of idealised secondary structure packing are approximated well by simple models (79–82).

Cohen and coworkers considered rigid units of secondary structure and combinatorially generated all possible three-dimensional arrangements between them (80, 82, 83). Chain connectivity, topological restrictions, and steric constraints were

applied to eliminate incorrect or impossible structures. The number of resultant structures was reduced to a manageable set (on the order of 10–100), and the correct structure was always present within the set of solutions. There was no way to, differentiate the correct answer from the other possibilities, however. Indeed, Novotný and coworkers showed that it is difficult to correctly identify misfolded structures by standard available criteria, including molecular mechanics potential energy, minimization of surface area, surface charge density, atomic packing, and Coulomb electrostatic energy (84).

When making decisions based on empirical examples, the choices that are always correct are easy to identify, and the choices that are always incorrect are easy to eliminate. But knowledge-based modeling methods share a common problem: there will always be choices which are only occasionally correct, and the likelihood of consistently identifying these uncommon events is small. These unlikely but occasionally correct choices will never be picked, even when they are actually correct. Examples of this dilemma are seen in the combinatorially modeled structures of Interleukin-2 and Human Growth Hormone (85–87). In these cases, models for three-dimensional structures were generated prior to their experimental determination. Both structures were predicted correctly to adopt a four-helix bundle motif, and the regions of α -helix were assigned fairly well. However, topological arguments resulted in the rejection of both correct tertiary structures. The correct left-handed α -helical bundle fold was rejected for Interleukin-2, following the argument of Weber and Salemme that right-handed structures were favored (88). The correct topology of Human Growth Hormone was rejected as too convoluted and kinetically unlikely (88). A subsequent review of the structures of four-helix bundle proteins reveals the errors of both of these assumptions (89).

In general, homology based modeling works well to predict gross structural features, but problems in exact alignment of more distantly related sequences remain. Although Pearl and Taylor's model of the HIV-1 protease based on a multiple sequence alignment of nonviral aspartic proteases (90) was correct in overall fold when compared to the crystal structure (91), certain regions were misaligned by one or two residues, and the model had a 2.6 Å rms deviation for 87 of the 99 C α atom positions. In contrast, a model based on a more closely related protease from Rous Sarcoma Virus (92) gave an rms deviation of 1.3 Å for 82 C α positions. In these models and others, the small shifts and reorientations of secondary structural elements relative to one another have proven difficult to model from homologous starting structures. Optimization and refinement of the model structures are necessary, but can introduce new problems into the structure. Overall compaction of the structure and distortion of the surface side chains can take place, especially if the protein is simulated in the gas phase (34, 93). Energy minimization can mask defects in a structure; molecular mechanics energy is not an adequate indicator of conformational correctness (84). In addition, energy minimization can disrupt side chain packing (63). Sequence misalignments of one or two residues are common, which can lead to the misalignment of β -strands or the incorrect facing of α -helices. Energy minimization or molecular dynamics cannot overcome the barriers necessary to reorganize such incorrect secondary structure packing, since it would be necessary to break hydrogen bonds or reorient entire helices to correct the misalignments.

The problems previously described for molecular mechanics and dynamics apply here as well. In particular, the inclusion of solvent, although rarely carried out for

computational reasons, may be crucial in the evaluation of the free energy and stability of models. Another concern is the power of MD algorithms to search conformational space. A stable structure will not travel far in phase space from its initial starting conformation in the course of a short trajectory. It may therefore be unable to cross energetic barriers to the new overall conformation that is the presumed energy minimum of the force field for the modeled amino acid sequence. It is also important to remember that molecular mechanics calculations measure the potential energy of a system, not the free energy. The calculated low energy configuration of a system will not necessarily have the lowest free energy.

Semi-Empirical and Polymer Models

It is possible to create simplified models that avoid atomic detail and instead rely on a more schematic residue-based view. A great deal is known from polymer theory about the behavior of these simplified models (94). Models can be constructed with a minimal amount of information, and tested to see if they exhibit the behavior of actual proteins. Fundamental questions can then be raised about protein stability. These include the relative importance of various constraints and intramolecular interactions. In this way, qualitative insights into protein conformation and folding can be gained. Since these models are simplified, many of the terms in the potential function do not correspond directly to actual energies, but instead are parameterized empirically to produce observed properties of interest. Skolnick and Kolinski have recently reviewed these topics (95).

One way of limiting the conformational space available to a protein is to confine a model polypeptide to a lattice. In doing so, unrealistic distortions are imposed on protein structure. However, lattice models offer the possibility to enumerate the entire conformation space available to a polymer chain. A detailed atomic picture is not typically employed with lattice models. However, a variety of lattices of increasing complexity facilitate more detailed chain representations. A trade-off exists between the detail of the models and the ability to evaluate conformational alternatives exhaustively.

Gō and coworkers have examined protein folding using Monte Carlo algorithms on two- and three-dimensional square and cubic lattices (96). They observe the cooperative “all-or-nothing” transitions to and from folded states and conclude this is a result of the specificity of interactions between residues separated along the chain. In these simulations an artificial potential function was created where interactions were ignored between residues that were not neighbors in the native state. Skolnick and Kolinski point out that the cooperativity of folding observed may be a result of intrinsic bias of considering only native interactions in this model (95).

Lau and Dill have also investigated the statistical mechanics of folding for simplified protein models on two-dimensional lattices (97, 98). They explored both conformation space (the set of all possible conformations) and sequence space (the set of all possible sequences), and concluded that many sequences have stable, compact, native-like structures. Another conclusion of these studies was that sequences tended to form a single unique structure, even with only two types of residues (hydrophobic and polar). This tendency increased with chain length. Moreover, one or two mutations in these sequences did not greatly destabilize most folded states.

Skolnick and coworkers have carried out a series of Monte Carlo calculations of an α -carbon chain on a tetrahedral lattice (95). This type of lattice allows for slightly more realistic three-dimensional conformations to be represented. Their calculations allowed for spontaneous folding of protein-like β -barrel and α -helix bundle structures, given parameters to simulate appropriate conformational preferences. They have also constructed a more realistic lattice model in which the chain can travel on a cubic lattice two units in one direction and one unit in another direction, similar to the move of a knight on a chessboard. Using this 210 "knight's move" on a cubic lattice, more accurate representations of actual protein topological features can be simulated (J. Skolnick, personal communication).

These lattice models seem to reproduce qualitative features of proteins, and the general conclusions of these studies appear valid. However, these simplifications may create quantitative discrepancies between structural results obtained using lattice and continuum representations of a protein. Simple cubic lattices are capable of representing virtual ($C^\alpha-C^\alpha-C^\alpha-C^\alpha$) dihedral angles of 0, 90, 180 and 270 degrees with virtual ($C^\alpha-C^\alpha-C^\alpha$) bond angles of 90 or 180 degrees. This sparse representation is inadequate to accurately sample the conformations available to proteins. In figure 3 we have illustrated the ability of two more complex lattice models to replicate actual properties of polypeptide chains. The knight's move cubic lattice is able to more accurately reproduce virtual bond and dihedral angles. Covell and Jernigan represented proteins on a face-centered cubic lattice, where additional nodes were placed at the centers of each of the cubic faces (99). This lattice also allows the actual structure of a protein to be well represented. It was possible to model C^α chains on this lattice with less than 1 Å rms deviation of C^α atoms from their positions in the crystal structure. This compares well with the 2.25 Å or greater rms deviation when using a simple cubic lattice. Note that in both the simple and knight's move cubic lattices, the internodal distance can be set to the 3.8 Å separation of C^α atoms in the *trans* peptide bond. In the face-centered cubic lattice, the greater conformational sampling gained is at the expense of introducing variable distances between C^α atoms, ranging from 2.69 to 4.65 Å. In these lattices the additional conformational freedom is given by increased connectivity. Each node in the knight's move and face-centered cubic lattice is connected to 24 and 42 others, respectively, compared to 6 for the cubic lattice. Therefore, it is no longer possible to exhaustively explore conformation space for chain lengths commonly seen in globular proteins. Even though these two lattices can sample the conformations of actual proteins well, as shown in figure 3, there is a much greater range of virtual angles available to the lattice representation than is found in actual proteins. Care must be taken in random walks and Monte Carlo calculations on such lattices that this region of conformational space is avoided.

Most lattice methods rely on an extremely simple potential function, either a two-state interresidue contact energy corresponding to native/non-native contacts, or a three-state model, corresponding to hydrophobic-hydrophobic, hydrophilic-hydrophilic, and hydrophobic-hydrophilic interactions. The interaction of the twenty naturally occurring amino acids in real proteins are obviously more complex.

Gregoret and Cohen calculated pairwise interaction pseudo-energies for all combinations of amino acids based on the three-dimensional distribution of residues in proteins of known structure (63). Equilibrium arguments were used to transform these distributions into free energies of association. This empirical formulation avoids the dilemma of choosing an appropriate effective dielectric constant, or accurately

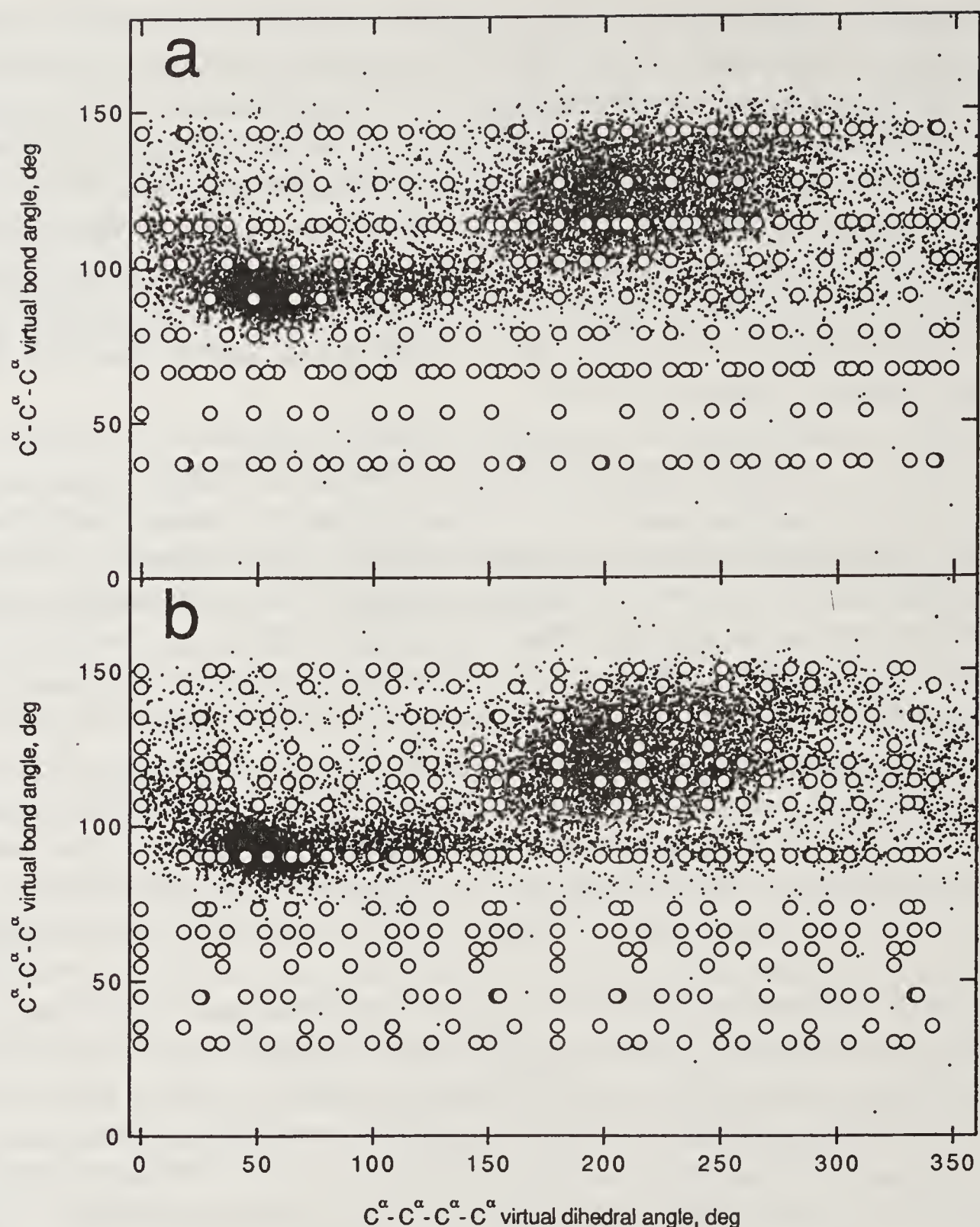


Figure 3. Distribution of virtual ($C^\alpha-C^\alpha-C^\alpha$) bond angles vs. virtual ($C^\alpha-C^\alpha-C^\alpha-C^\alpha$) dihedral angles using simplified representations of protein structure. The actual geometries of the 72 proteins listed in reference (63) are plotted as dots. (a) \circ , Available lattice geometries for a self-avoiding “knight’s walk” on a cubic lattice. Conformations with three colinear atoms are omitted. (b) Similar to (a) for a face-centered cubic lattice, with nodes at the center of each cubic face.

modeling protein-solvent and solvent-solvent interactions in the calculations. Empirical pair potentials intrinsically incorporate energetic terms such as London dispersion forces, charge-charge interactions, and hydrogen bonding. Constraints resulting from protein organization, such as the tendency for hydrophobics to cluster together, are included implicitly. The sum of pair potentials over all nearest spatial neighbors in the protein was helpful in differentiating between correct and incorrect models.

Similar pairwise potentials have been developed by Tanaka and Scheraga to investigate protein folding using a Monte Carlo algorithm (100). Narayana and Argos also calculated contact energies which they used in interaction correlation functions to investigate homology in distantly related sequences (101). Miyazawa and Jernigan

calculated similar contact energies in a somewhat more sophisticated manner (102). All of these pairwise interaction matrices show similar overall properties: hydrophobic residues tend to associate together. Oppositely charged residues show a similar behavior.

Levitt took an opposite approach in calculating residue–residue interaction energies, and based his calculations on theoretical considerations rather than empirical distributions (103). He assumed that the average interaction between side chains would be the average of the calculated interaction over all possible spatial orientations of the two residues. This assumption is hampered by the fact that side chain structures favor certain interaction geometries in folded proteins (104). Backbone conformations also influence side chain dihedral angles (71).

With a residue-based model it becomes computationally feasible to carry out simulations of protein folding without using a lattice approximation. Early work by Levitt and Warshel and Kuntz *et al.* attempted to correctly fold BPTI from an extended conformation (105, 106). Levitt and Warshel used a C^α model with side chains represented by a single sphere located at its centroid. Force field terms were created through spatial averaging of all-atom models. They included side chain–side chain and backbone–backbone nonbonded interactions, backbone hydrogen bonding, and sidechain–solvent interactions as a function of approximate solvent accessibility. Conformations were explored using energy minimization. A normal mode thermalization procedure was employed to avoid local minima. Kuntz and coworkers used a more arbitrary method, incorporating functionally useful terms in their force field, without attempting to correlate these terms with the actual energetics of protein folding. A “beads-on-a-string” model was used, with virtual bonds connecting the C^β atoms of successive residues. The force included hydrophobic attraction terms as well as terms constraining hydrophobic and hydrophilic residues to certain target distances from the protein center. Simple minimization of this error function over cartesian coordinates yielded compact structures.

The approaches of Levitt and Warshel and Kuntz *et al.* generated structures which exhibited topological features reminiscent of the BPTI native structure. However, Hagler and Honig subsequently noted that several important features of the BPTI fold were not present, and that similar results could be obtained given simple models composed of polyalanine with glycine residues inserted at turns (107). Cohen and Sternberg showed that chains generated by random walks constrained to produce compact structures have rms deviations from the native structure comparable to the BPTI simulation results (108). This suggests that little true structural information was successfully incorporated into early simplified potential functions.

A more recent simulation using a simplified chain representation was performed by Wilson and Doniach (109). They attempted to fold crambin from a random conformation using a Monte Carlo simulated annealing algorithm. Their model consisted of a polypeptide backbone with side chains represented by a single atom at their centroid. Empirical potentials were calculated for combined ϕ – ψ energy and residue i to residue $i + 4$ interactions. Side chain interaction energies were calculated from pairwise radial distribution functions. Conformations were generated by perturbing the backbone dihedral angles according to a Metropolis algorithm. Qualitative agreement with the overall fold of crambin was achieved, including the correct formation of secondary structure. Wilson and Doniach also addressed the crucial questions of convergence and correct energy minimum. The simulated

structures did not converge to a single conformation, although the averaged distance matrix was closer to the native distance matrix than any single structure. Thus the simulation did not always converge to the global energy minimum. Non-native structures were found with lower energies than the native structure. This indicates that the force field was suboptimally parameterized, since the native structure did not correspond to the global energy minimum.

The relative success of these crude simulations suggest that simple models using empirical parameters may be able to correctly represent globular protein structure. A correctly parameterized simplified force should make reasonable, physically appropriate assumptions, and should have a global energy minimum corresponding to the native structures of proteins. Initial parameters could be taken from Brookhaven Protein Data Bank averages, and be further refined using Consistent Force Field methods (110, 111).

A simplified force field could be used in conjunction with any conformational search technique, since the generation of test structures is much faster than the subsequent evaluation of their relative energies. When used in conjunction with molecular dynamics or Monte Carlo simulations, much longer calculations would be possible due to the decreased time needed to evaluate pairwise nonbonded interactions. In addition, since the potential energy terms in a simplified force field would correspond to softer degrees of freedom, longer step sizes of integration could be taken through phase space. Since many of the minor energy minima of all-atom representations should be smoothed out with a simplified force field, the radius of convergence should be larger, and it may be possible to refine approximate structures and models into the correct area of conformational space, where they could then be further optimized using a more detailed force field such as AMBER (93).

Current prediction techniques allow a detailed and reasonable structure to be generated only for small or constrained systems. Five- or ten-residue protein loops or molecules such as enkephalin represent the limit of current techniques. A larger database of protein conformational motifs may increase the success of prediction (112), but brute force computational approaches will remain infeasible for some time. The incorporation of *a priori* knowledge, whether at an atomic level or through abstraction at a higher level or organization, appears to be the best means available to understand and predict protein conformation.

Acknowledgements

We would like to thank Alison Howard, Lydia Gregoret, Scott Presnell, and Christine Ring for their helpful suggestions in the preparation of this manuscript. This work was supported by the National Institutes of Health (GM 39900), the Searle Scholars/Chicago Community Trust and the Defense Advanced Research Projects Agency (N00014-86-12-0757).

References

1. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. (1961) *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1309–1314.
2. Blundell, T. L. & Johnson, L. N. (1976) *Protein Crystallography* (Academic Press, New York).

3. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (J. Wiley & Sons, New York).
4. McCammon, J. A. & Harvey, S. C. (1987) *Dynamics of Proteins and Nucleic Acids* (Cambridge University Press, New York).
5. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963) *J. Mol. Biol.* **7**, 95–97.
6. Brooks III, C. L., Karplus, M. & Pettitt, B. M. (1988) *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics* (J. Wiley & Sons, New York).
7. Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E. & Thornton, J. M. (1987) *Nature (London)* **326**, 347–352.
8. Howard, A. E. & Kollman, P. A. (1988) *J. Med. Chem.* **31**, 1675–1679.
9. Burkert, U. & Allinger, N. L. (1982) *Molecular Mechanics* (American Chemical Society, Washington, D. C.).
10. Loncharich, R. J. & Brooks, B. R. (1989) *Proteins Struct., Funct. Genet.* **6**, 32–45.
11. Brooks III, C. L., Pettitt, B. M. & Karplus, M. (1985) *J. Chem. Phys.* **83**, 5897–5908.
12. van Gunsteren, W. F. & Karplus, M. (1982) *Macromolecules* **15**, 1528–1544.
13. Wendolowski, J. J., Matthew, J. B., Weber, P. C. & Salemme, F. R. (1987) *Science* **238**, 794–797.
14. Koehler, J. E. H., Saenger, W. & van Gunsteren, W. F. (1988) *J. Biomol. Struct. Dyn.* **6**, 181–198.
15. Makinen, M. W., Troyer, J. M., van der Werff, H., Berendsen, H. J. & van Gunsteren, W. F. (1989) *J. Mol. Biol.* **207**, 201–216.
16. Scheek, R. M., van Gunsteren, W. F. & Kaptein, R. (1989) *Methods Enzymol.* **177**, 204–218.
17. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989) *Science* **245**, 635–637.
18. Clore, G. M., Nilges, M., Brünger, A. T., Karplus, M. & Gronenborn, A. M. (1987) *FEBS Lett.* **213**, 269–277.
19. Brünger, A. T., Kuriyan, J. & Karplus, M. (1987) *Science* **235**, 458–460.
20. van Gunsteren, W. F. (1988) *Protein Eng.* **2**, 5–13.
21. Warshel, A. & Sussman, F. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3806–3810.
22. Bash, P. A., Singh, U. C., Brown, F. K., Langridge, R. & Kollman, P. A. (1987) *Science* **235**, 574–576.
23. Beveridge, D. L. & DiCapua, F. M. (1989) *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431–492.
24. Reid, R. H., Hooper, C. A. & Brooks, B. R. (1989) *Biopolymers* **28**, 525–530.
25. Saunderson, M., Houk, K. N., Wu, Y.-D., Still, W. C., Lipton, M., Change, G. & Guida, W. C. (1990) *J. Am. Chem. Soc.* **112**, 1419–1427.
26. Kuriyan, J., Brünger, A. T., Karplus, M. & Hendrickson, W. A. (1989) *Acta Cryst.* **A45**, 396–409.
27. Wilson, S. R. & Cui, W. (1990) *Biopolymers* **29**, 225–235.
28. Noguti, T. & Go, N. (1985) *Biopolymers* **24**, 527–546.
29. Noguti, T. & Go, N. (1989) *Proteins Struct., Funct. Genet.* **5**, 97–138.
30. Northrup, S. H. & McCammon, J. A. (1980) *Biopolymers* **19**, 1001.
31. Noguti, T. & Go, N. (1982) *Nature (London)* **296**, 776–778.
32. Levy, R. M., Srinivasan, A. R., Olson, W. K. & McCammon, J. A. (1984) *Biopolymers* **23**, 1099–1112.
33. van Gunsteren, W. F. & Karplus, M. (1982) *Biochemistry* **21**, 2259–2273.
34. Brooks III, C. L. & Karplus, M. (1989) *J. Mol. Biol.* **208**, 159–181.
35. MacKerell, Jr., A. D., Nilsson, L., Rigler, R. & Saenger, W. (1988) *Biochemistry* **27**, 4547–4556.
36. Brünger, A. T., Brooks III, C. L. & Karplus, M. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8458–8462.
37. Jorgensen, W. L. & Gao, J. (1988) *J. Am. Chem. Soc.* **110**, 4212–4216.
38. Anderson, A. G. & Hermans, J. (1988) *Proteins Struct., Funct. Genet.* **3**, 262–265.
39. Ghosh, I. & McCammon, J. A. (1987) *Biophys. J.* **51**, 637–641.
40. Harvey, S. C. (1989) *Proteins Struct., Funct. Genet.* **5**, 78–92.
41. Davis, M. E. & McCammon, J. A. (1990) *Chem. Rev.* **90**, 509–521.
42. Warshel, A. (1979) *J. Phys. Chem.* **83**, 1640–1652.
43. Howard, A. E., Singh, U. C., Billeter, M. & Kollman, P. A. (1988) *J. Am. Chem. Soc.* **110**, 6984–6991.
44. Matthew, J. B. (1985) *Annu. Rev. Biophys. Biophys. Chem.* **14**, 387–417.
45. Warshel, A. & Russell, S. T. (1984) *Q. Rev. Biophys.* **17**, 283–422.
46. Rees, D. C. (1980) *J. Mol. Biol.* **141**, 323–326.

47. Warshel, A., Russell, S. T. & Churg, A. K. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4785–4789.
48. Gilson, M. K. & Honig, B. H. (1986) *Biopolymers* **25**, 2097–2119.
49. Wendolowski, J. J. & Matthew, J. B. (1989) *Proteins Struct., Funct. Genet.* **5**, 313–321.
50. Creighton, T. E. (1978) *Prog. Biophys. Mol. Biol.* **33**, 231–297.
51. Kauzmann, W. (1959) *Adv. Protein Chem.* **14**, 1–63.
52. Dill, K. A. (1990) *Biochemistry* **29**, 7133–7155.
53. Cornette, J. L., Cease, K. B., Margalit, H., Spouge, J. L., Berzofsky, J. A. & DeLisi, C. (1987) *J. Mol. Biol.* **195**, 659–685.
54. Chothia, C. (1974) *Nature (London)* **254**, 304.
55. Eisenberg, D. & McLachlan, A. D. (1986) *Nature (London)* **319**, 199–203.
56. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379–400.
57. Connolly, M. L. (1986) *Int. J. Pept. Protein Res.* **28**, 360–363.
58. Wodak, S. J. & Janin, J. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1736–1740.
59. Richmond, T. J. (1984) *J. Mol. Biol.* **178**, 63–89.
60. Kang, Y. K., Némethy, G. & Scheraga, H. (1987) *J. Phys. Chem.* **91**, 4105–4109.
61. Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.
62. Rashin, A. A. (1984) *Biopolymers* **23**, 1605–1620.
63. Gregoret, L. M. & Cohen, F. E. (1990) *J. Mol. Biol.* **211**, 959–974.
64. Greer, J. (1985) *Science* **228**, 1055–1060.
65. Bashford, D., Chothia, C. & Lesk, A. M. (1987) *J. Mol. Biol.* **196**, 198–216.
66. Chothia, C. & Lesk, A. M. (1982) *J. Mol. Biol.* **160**, 309–323.
67. Lesk, A. M. & Chothia, C. (1982) *J. Mol. Biol.* **160**, 325–342.
68. Jones, T. A. & Thirup, S. (1986) *EMBO J.* **5**, 819–822.
69. Unger, R., Harel, D., Wherland, S. & Sussman, J. L. (1989) *Proteins Struct., Funct. Genet.* **5**, 355–373.
70. Ponder, J. W. & Richards, F. M. (1987) *J. Mol. Biol.* **193**, 775–791.
71. McGregor, M. J., Islam, S. A. & Sternberg, M. J. E. (1987) *J. Mol. Biol.* **198**, 295–310.
72. Dudek, M. J. & Scheraga, H. A. (1990) *J. Comput. Chem.* **11**, 121–151.
73. Vasquez, M. & Scheraga, H. A. (1988) *J. Biomol. Struct. Dyn.* **5**, 705–755.
74. Moulton, J. & James, M. N. G. (1986) *Proteins Struct., Funct. Genet.* **1**, 146–163.
75. Shenkin, P. S., Yarmush, D. L., Fine, R. M., Wang, H. & Levinthal, C. (1987) *Biopolymers* **26**, 2053–2085.
76. Bruccoleri, R. E. & Karplus, M. (1987) *Biopolymers* **26**, 137–168.
77. Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Scheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M. & Poljak, R. J. (1989) *Nature (London)* **343**, 877–883.
78. Chothia, C. (1984) *Annu. Rev. Biochem.* **53**, 537–572.
79. Richmond, T. J. & Richards, F. M. (1978) *J. Mol. Biol.* **119**, 537–555.
80. Cohen, F. E., Richmond, T. J. & Richards, F. M. (1979) *J. Mol. Biol.* **132**, 275–288.
81. Cohen, F. E., Sternberg, M. J. E. & Taylor, W. R. (1981) *J. Mol. Biol.* **148**, 253–272.
82. Cohen, F. E., Sternberg, M. J. E. & Taylor, W. R. (1982) *J. Mol. Biol.* **156**, 821–862.
83. Cohen, F. E., Sternberg, M. J. E. & Taylor, W. R. (1980) *Nature (London)* **285**, 378–382.
84. Novotný, J., Rashin, A. A. & Bruccoleri, R. E. (1988) *Proteins Struct., Funct. Genet.* **4**, 19–30.
85. Cohen, F. E., Kosen, P. A., Kuntz, I. D., Epstein, L., Ciardelli, T. & Smith, K. A. (1986) *Science* **234**, 349–352.
86. Cohen, F. E. & Kuntz, I. D. (1987) *Proteins Struct., Funct. Genet.* **2**, 162–166.
87. Cohen, F. E., Gregoret, L., Presnell, S. R. & Kuntz, I. D. (1989) in *Computer-Assisted Modeling of Receptor-Ligand Interactions: Theoretical Aspects and Applications to Drug Design*, eds. Rein, R. & Golombek, A. (Alan R. Liss, New York), pp. 75–85.
88. Weber, P. C. & Salemme, F. R. (1982) *Nature (London)* **287**, 82–84.
89. Presnell, S. R. & Cohen, F. E. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6592–6596.
90. Pearl, L. H. & Taylor, W. R. (1987) *Nature (London)* **329**, 351–354.
91. Wlodawer, A., Miller, M., Jaskolski, M., Sathynarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. & Kent, S. B. H. (1989) *Science* **245**, 616–621.
92. Weber, I. T. (1990) *Proteins Struct., Funct. Genet.* **7**, 172–184.

93. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, Jr., S. & Weiner, P. (1984) *J. Am. Chem. Soc.* **106**, 765–784.
94. Flory, P. J. (1989) *Statistical Mechanics of Chain Molecules* (Oxford University Press, New York).
95. Skolnick, J. & Kolinski, A. (1989) *Annu. Rev. Phys. Chem.* **40**, 207–235.
96. Go, N. & Taketomi, H. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 559–563.
97. Lau, K. F. & Dill, K. A. (1989) *Macromolecules* **22**, 3986–3997.
98. Lau, K. F. & Dill, K. A. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 638–642.
99. Covell, D. G. & Jernigan, R. L. (1990) *Biochemistry* **29**, 3287–3294.
100. Tanaka, S. & Scheraga, H. A. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3802–3806.
101. Narayana, S. V. L. & Argos, P. (1984) *Int. J. Pept. Protein Res.* **24**, 25–39.
102. Miyazawa, S. & Jernigan, R. L. (1985) *Macromolecules* **18**, 534–552.
103. Levitt, M. (1976) *J. Mol. Biol.* **104**, 59–107.
104. Singh, J. & Thornton, J. M. (1990) *J. Mol. Biol.* **211**, 595–615.
105. Levitt, M. & Warshel, A. (1975) *Nature (London)* **253**, 694–698.
106. Kuntz, I. D., Crippen, G. M., Kollman, P. A. & Kimelman, D. (1976) *J. Mol. Biol.* **106**, 983–994.
107. Hagler, A. T. & Honig, B. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 554–558.
108. Cohen, F. E. & Sternberg, M. J. E. (1980) *J. Mol. Biol.* **138**, 321–333.
109. Wilson, C. & Doniach, S. (1989) *Proteins Struct., Funct. Genet.* **6**, 193–209.
110. Lifson, S. & Warshel, A. (1968) *J. Chem. Phys.* **49**, 5116–5129.
111. Niketić, S. R. & Rasmussen, K. (1977) *The Consistent Force Field: A Documentation* (Springer-Verlag, New York).
112. Wodak, S. J. & Rooman, M. J. (1988) *Nature (London)* **335**, 45–49.

Conformations for Strand Entry into Parallel β -Sheets

Cyrus Chothia¹ and Arthur M. Lesk²

¹Cambridge Centre for Protein Engineering, and
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

²Department of Haematology, University of Cambridge Clinical School,
Addenbrookes' Hospital, Cambridge CB2 2QL, UK

Introduction

The identification of the low-energy conformations available to polypeptides (1) has had a profound influence on our understanding of protein structure. Over the past thirty years, these conformations have been used continually to define, analyse and explain important features of protein structure. In this paper, we describe a new application: the conformations involved in the entry of strands into parallel β -sheets.

Our results are derived from analysis of seven proteins that contain parallel β -sheets, the structures of which have been refined at high resolution—between 1.2 and 2.0 Å (table 1). We show that, for more than three-quarters of the strands of these proteins, the residues at the point at which they enter the β -sheet have one of two well-defined conformations. The few exceptions are strands that contain glycine in the entry region or those that protrude from the β -sheet structure. We first describe the general features of the two conformations and then discuss individual cases.

The Two Conformations for Strand Entry Into Parallel β -Sheets

In protein structures, the polypeptide chain usually runs back and forth across the protein, with each run consisting of an α -helix or a strand in a β -sheet. The secondary structures are linked by turns in the polypeptide chain. These features might suggest that there would be two ways for a polypeptide chain to turn into the extended conformation of a strand. One would be through a conformation near that of the α -helix; this would have the effect of kinking the chain. The other would be through residues in the β region; this would produce a local coiling of the chain.

Inspection of well-determined protein structures shows that both these expected alternatives occur commonly, and furthermore, that rather precise conformations within these general regions are observed. We shall call the two conformations A and B. Both conformations involve two adjacent residues; the first residue in the strand to form β -sheet hydrogen bonds and the residue immediately prior to this one.

Table 1. Some well-determined protein structures containing parallel β -sheets.

Protein	Structure determination		
	Resolution (\AA)	R factor (%)	Reference
Carboxypeptidase A (CPA)	1.5		Rees <i>et al.</i> (2)
Dihydrofolate reductase (DHFR)	1.7	15	Bolin <i>et al.</i> (3)
Glucose binding protein (GBP)	1.9	15	Vyas <i>et al.</i> (4)
Glyceraldehyde-3-phosphate dehydrogenase (GPD)	1.8	18	Skarzynski <i>et al.</i> (5)
Muconate lactonizing enzyme (MLE)	1.8	19	Goldman <i>et al.</i> (6)
N-(5'-phosphoribosyl) anthranilate isomerase-indole-3-glycerol-phosphate synthase (PRAI-IGPS)	2.0	17	Priestle <i>et al.</i> (7)
Subtilisin (SUB)	1.2	18	Bode <i>et al.</i> (8)

Figure 1 shows schematic pictures of these two conformations and the relation of the residues to the β -sheet.

In conformation A the first sheet residue has a conformation close to $\phi = -95^\circ, \psi = -25^\circ$, which is near that of an α -helix ($\phi = -65^\circ, \psi = -40^\circ$). The residues adjacent in the sequence have conformations in the β region (figure 1). This has the effect of producing a kink in the polypeptide chain (figure 2). In conformation B the two residues have conformations at points in the β region along the line $\phi + \psi \simeq +50^\circ$ (figure 1). Such values produce a local coiling of the strand, as shown in figure 3 (ref. 9, and references therein).

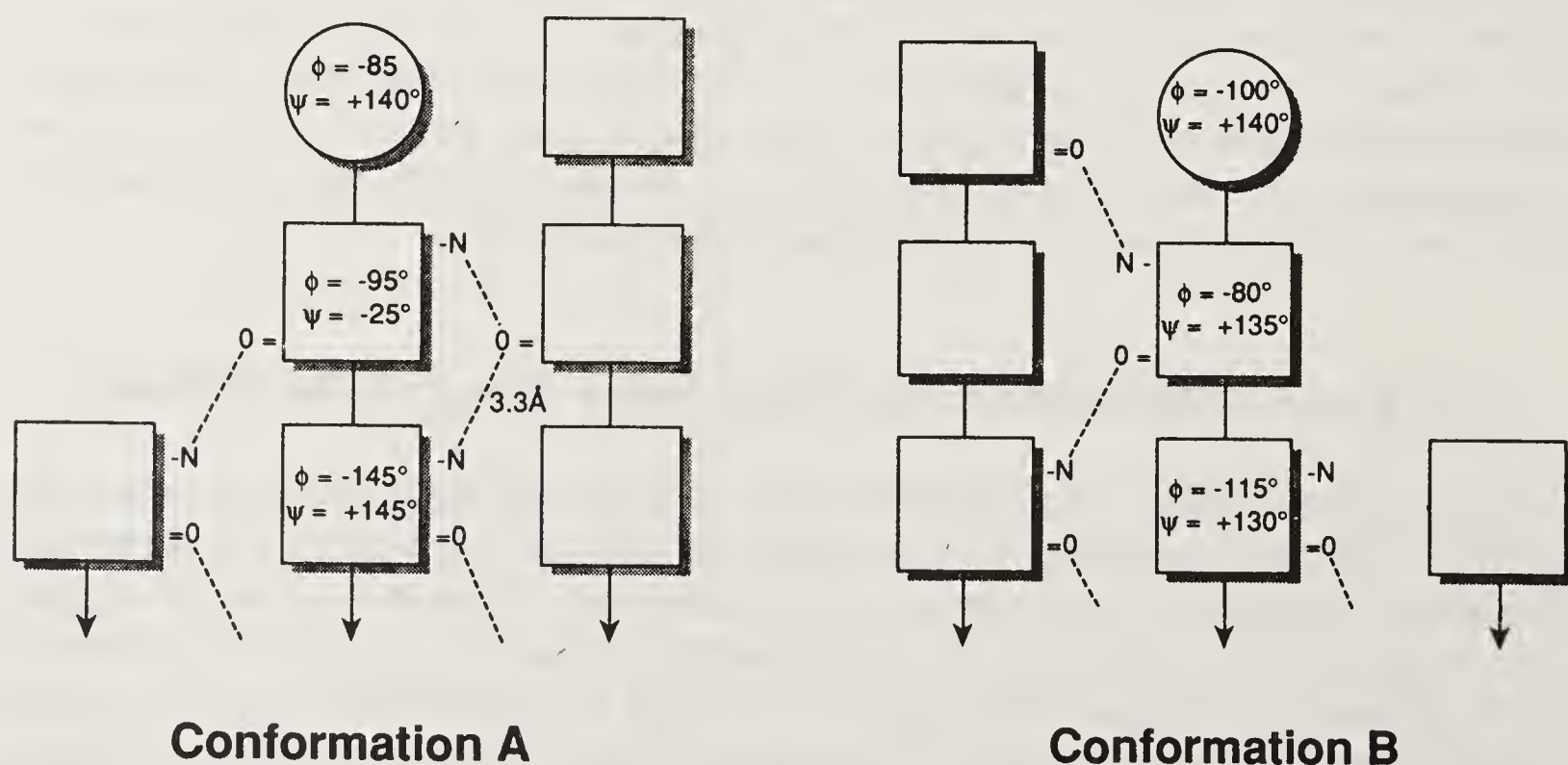


Figure 1. Conformations A and B for strand entry into β -sheets. Squares indicate residues forming hydrogen bonds within the β -sheet and circles indicate residues outside the β -sheet. The ϕ, ψ values shown here are the mean values given in table 2, rounded to the nearest 5° .

Not all of the hydrogen bonds shown are formed in every example of such conformations because of the absence or distortion of residues in the neighbouring strands (see figures 4–6).

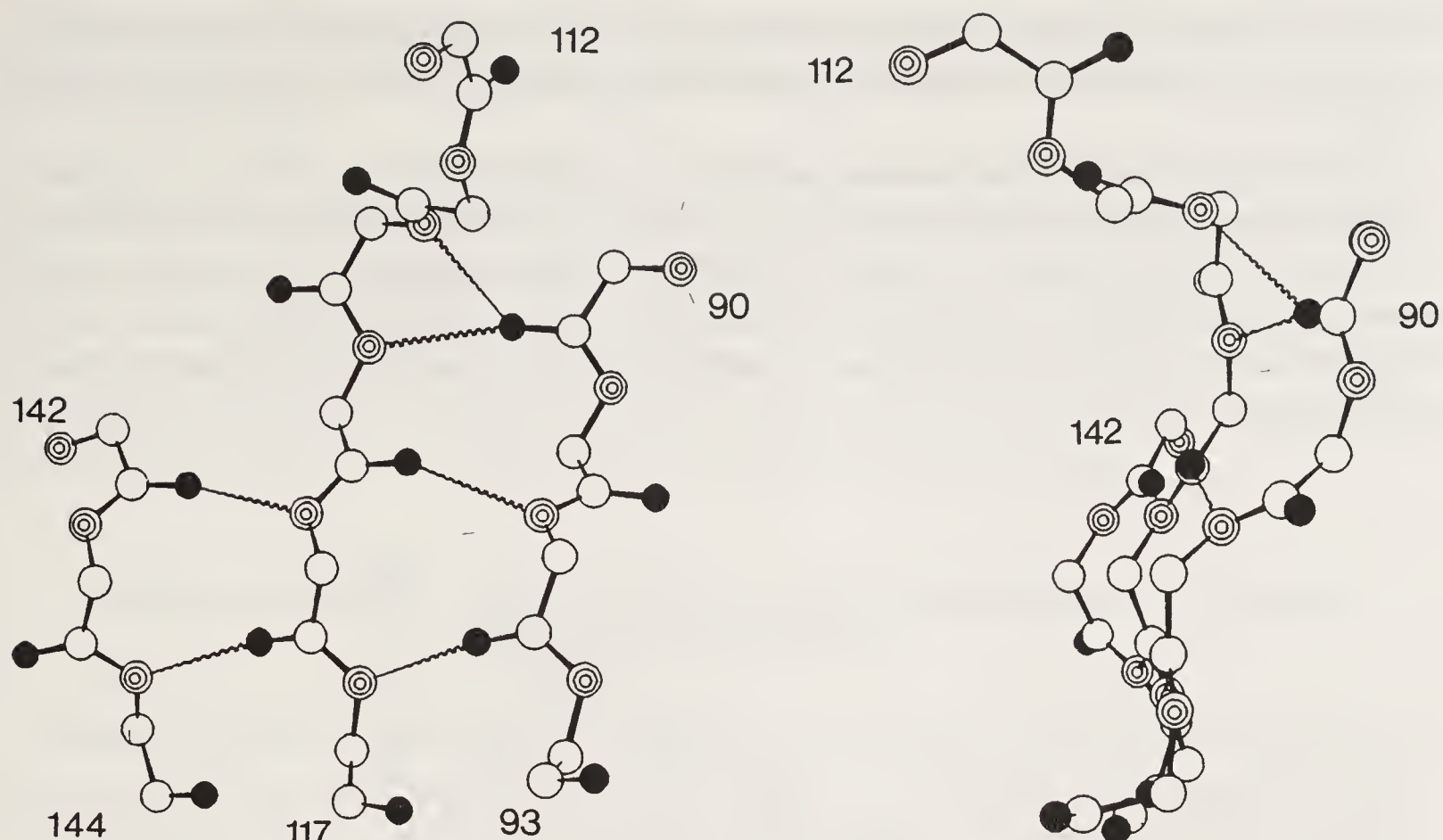


Figure 2. Two orthogonal views of an example of conformation A from glyceraldehyde-3-phosphate dehydrogenase. Residue 114 has $\phi = -94$, $\psi = -33$.

In conformation A, the helical conformation of the first sheet residue means that its main chain NH and CO groups cannot form hydrogen bonds on the same side of the strand but only on opposite sides. Thus, the NH groups of the first and second residues often hydrogen bond to the same carbonyl oxygen in the adjacent strand (see figure 1). These hydrogen bonds, not normally found in parallel β -sheets, are the

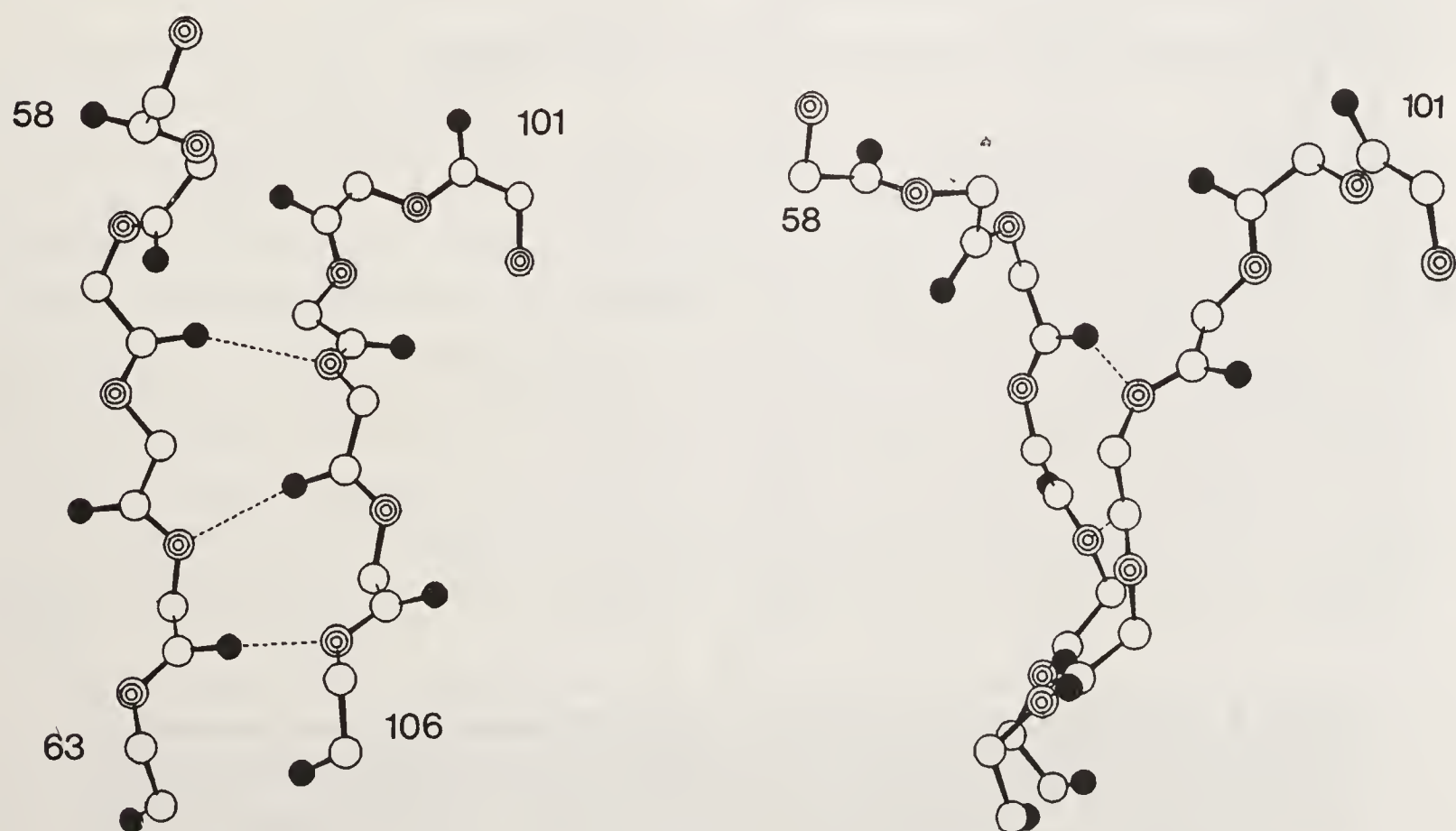


Figure 3. Two adjacent examples of conformation B from carboxypeptidase A. Residue 60 has $\phi = -44$, $\psi = +137$ and residue 104 has $\phi = -99$, $\psi = +155$.

most distinctive feature of the A conformation. The hydrogen bonds formed by residues in the B conformation are those that normally occur in parallel β -sheets (figure 1).

If the (ϕ, ψ) values of the residues that form conformations A and B are compared, it will be seen that the major difference is in the value of the angle ψ of the first sheet residue: -25° in the case of A and $+135^\circ$ in the case of B; the other torsion angles show only small differences (figure 1). The effect of this large localized difference in conformational angle is that chains in conformations A and B come from opposite directions.

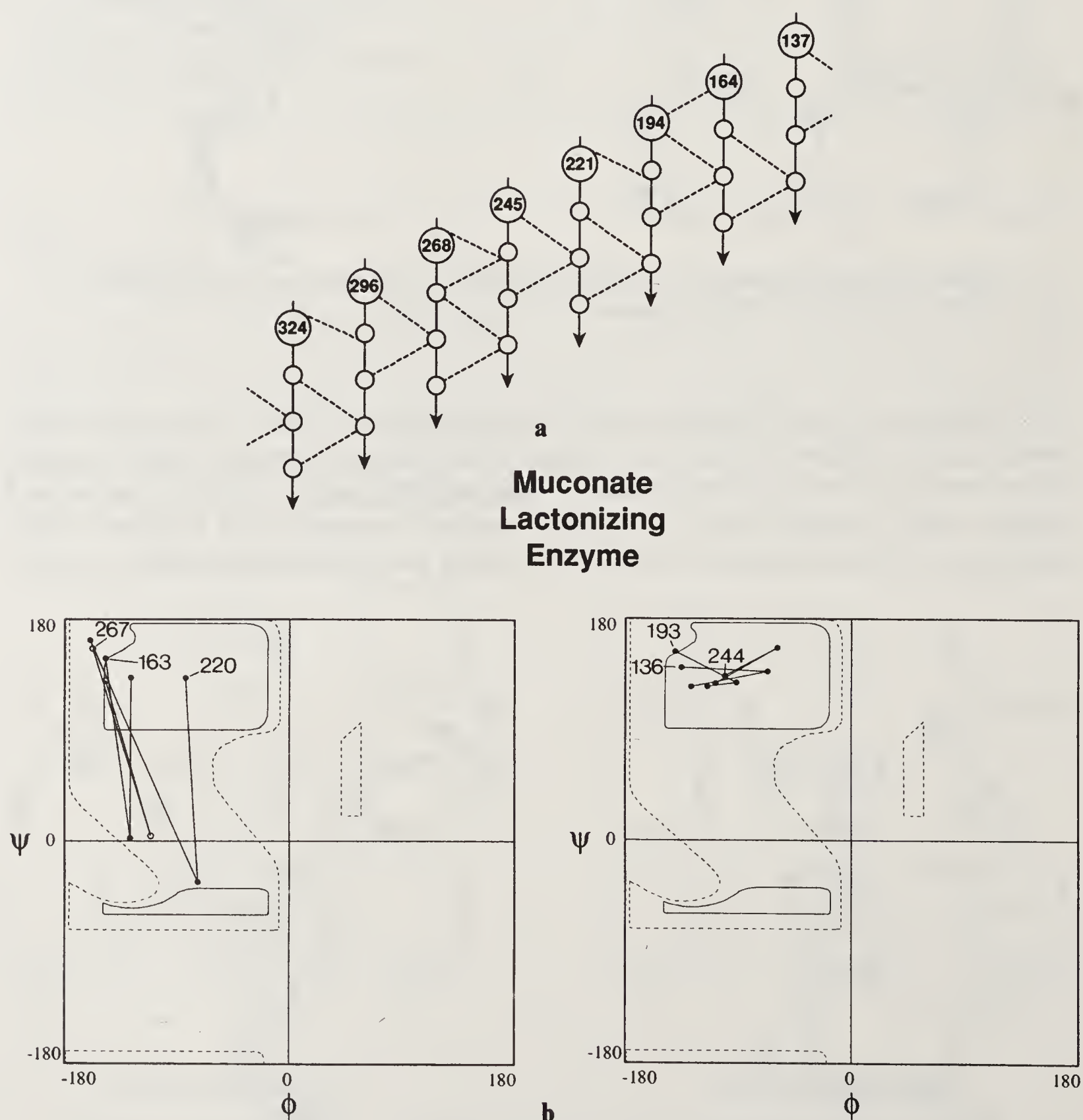


Figure 4. Conformations A and B in the α/β -barrel protein muconate lactonizing enzyme. (a) the N-terminal regions of the parallel β -sheet. (b) Ramachandran plots for the residues in conformations A (left) and B (right). We show the ϕ, ψ values of three adjacent residues: the first is outside the β -sheet structure (this residue bears a number in the diagram) and the other two are within the sheet, as shown in part (a).

This can be stated in more general terms: the residues in conformation B produce a local right-handed coiling of the polypeptide chain; those in conformation A produce a left-handed kink.

In conformation A, the side chain of the residue outside the β -sheet points straight over the strand and is therefore usually buried by the rest of the structure. Accordingly, the side chains at this site are generally hydrophobic or neutral. This is true for 23 of the 24 cases discussed below. In conformation B the position of this side chain is not so precisely located, and the structural constraints are much less stringent. Although hydrophobic or neutral residues often occur, long hydrophilic residues that can reach the solvent, e.g., arginine and glutamine, are also found.

Conformations of Strands Entering Parallel β -Sheets in Known Protein Structures

Parallel β -sheets occur mainly in α/β and α/β -barrel proteins. We examined the structures of the parallel β -sheets in the examples of both classes listed in table 1. These seven proteins have nine β -sheets, the strands of which are parallel, with only a few exceptions. In figures 4 to 6 we show the structures of the N-terminal regions of the seven fully-parallel sheets. In all, the structures contain 59 strands. It will be seen that a trace of the conformation produced by connecting the points that correspond to successive residues in a "Ramachandran plot" gives a perspicuous representation of the conformation of the region.

In table 2 we list the (ϕ, ψ) values of the residues involved in entry of the strands into the β -sheets.

Discussion

The data demonstrate clearly that peptides entering strands of parallel β -sheets have two preferred conformations.

The seven proteins we studied contain a total of 59 strands in parallel β -sheets. At their N-termini, 47 of these strands show one of the two conformations we have described and illustrated: there are 24 of type A and 23 of type B. Of the others, one class of exceptions involves edge strands of open sheets; that is, sheets that are not closed into barrels. Examples can be seen in residues 32 and 103 of glucose-binding protein (figure 6a), and residue 23 in glyceraldehyde-3-phosphate dehydrogenase (figure 6b). These residues are at the corners of the sheets. Another class of exceptions involves glycine residues, to which the usual restrictions to preferred α_R or β conformations do not apply. In one other exceptional case, residue 25 in subtilisin (figure 5), the strand is longer than its neighbours in the sheet and extends out beyond them.

In some cases the structural reason for adopting either of these conformations can be seen in the overall architectural constraints of the protein fold: Recall that the A and B conformations result in the strand entering the sheet from opposite sides. The α/β barrel proteins are cases in point. This folding pattern contains eight units of β -strand followed by α -helix; the strands close to form a cylindrical eight-stranded β sheet with the helices packed on the outside of the sheet. In the two α/β barrel

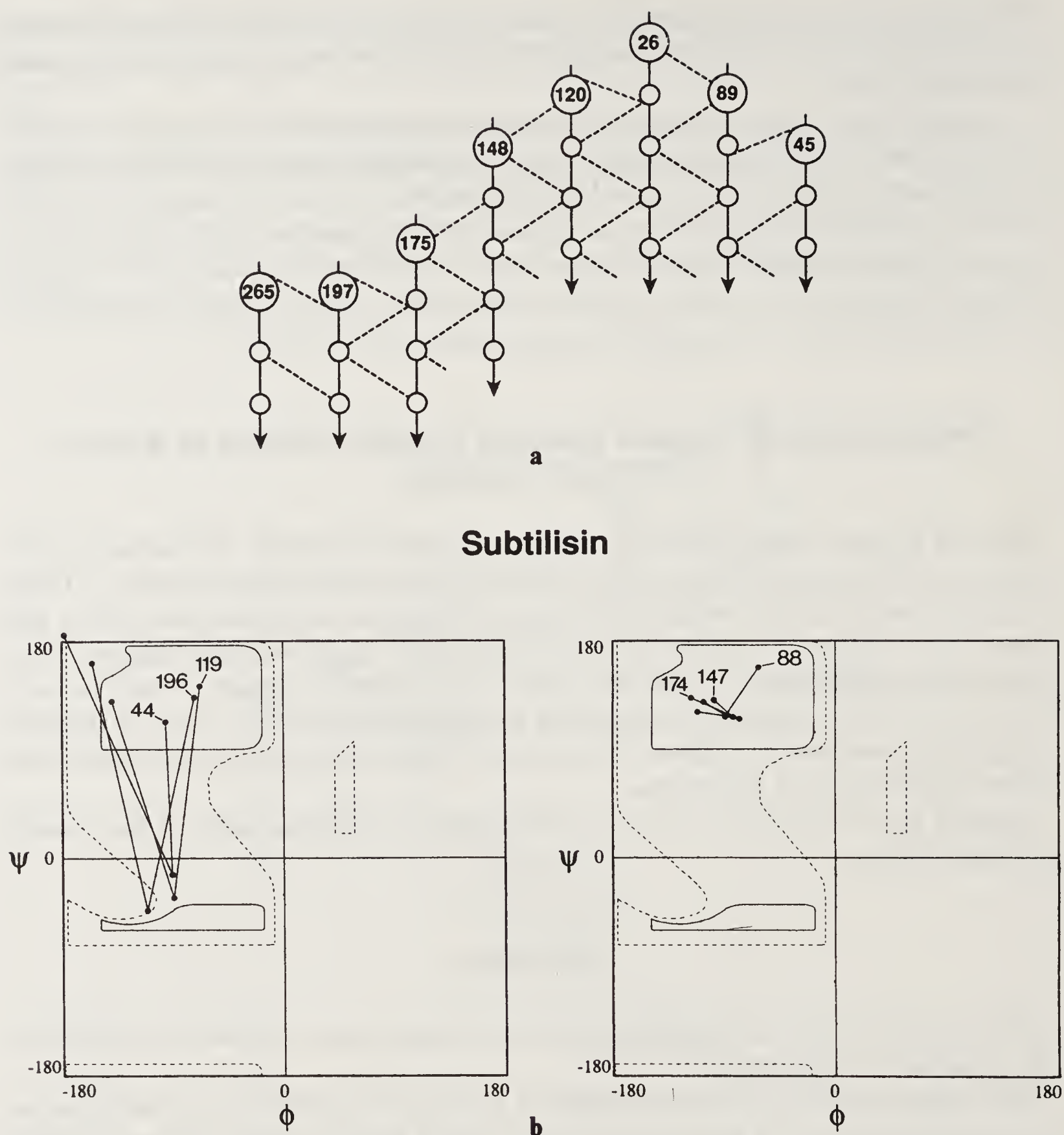


Figure 5. Conformations A and B in the α/β protein subtilisin. (a) the N-terminal regions of the parallel β -sheet. (b) Ramachandran plots for the residues in conformations A (left) and B (right); notation as described in the caption to figure 4(b).

proteins studied (muconate lactonizing enzyme and N-(5'-phosphoribosyl) anthranilate isomerase-indole-3-glycerol-phosphate synthase) there is an alternation of A and B conformations in successive strands. This is required to keep the helices on the outside of the sheet. To see this, note in figures 4a and 6c that adjacent strands are displaced by one residue along the sheet – a result of the tilt of the strands relative to the barrel axis. Because of this displacement, it is not possible to maintain the same relative position of the strand and the entry loop by keeping the same conformation (A or B) in the entry peptide, as occurs, for example in the strands beginning at residues 114 and 90 of glyceraldehyde-3-phosphate dehydrogenase. Instead, there must be an alternation of the entry conformation.

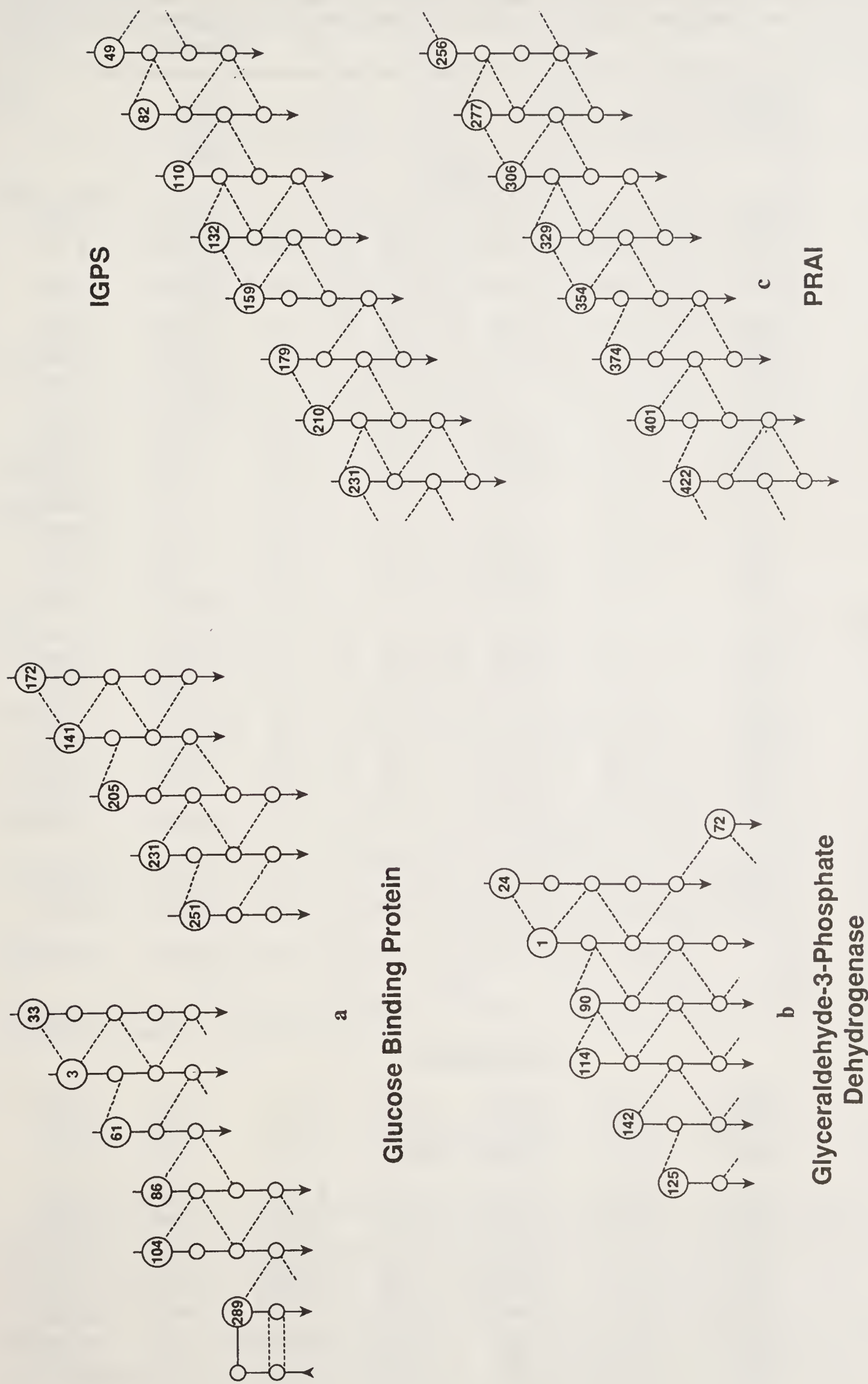


Figure 6. Diagrams of the N-terminal regions of the parallel β -sheets in (a) glucose binding protein (GBP), (b) glyceraldehyde-3-phosphate dehydrogenase (GPD) and (c) N-(5'-phosphoribosyl) anthranilate isomerase-indole-3-glycerol-phosphate synthase (PRAI-IGPS). Note that GBP and PRAI-IGPS contain two β -sheets.

Table 2. Conformation of residues at the entry of strands into parallel β -sheets.

We list here the (ϕ, ψ) values of three adjacent residues. The first, designated -1 , is outside the β -sheet structure, The next two, designated 1 and 2, are within the β -sheet. Refer to the diagrams of the β -sheets in figures 4–6.

Protein*	Initial Residue	ϕ_{-1}	ψ_{-1}	ϕ_1	ψ_1	ϕ_2	ψ_2
Conformation A							
CPA	189F	− 75	+ 130	− 107	− 25	− 132	+ 157
	263G	− 88	+ 104	− 68	− 40	− 91	+ 75
DHFR	110V	− 89	+ 141	− 113	− 22	− 136	+ 141
GBP	60V	− 66	+ 149	− 91	− 13	− 165	+ 153
	204I	− 65	+ 125	− 106	− 32	− 143	+ 144
	250L	− 113	+ 129	− 81	− 29	− 179	+ 158
GPD	89V	− 68	+ 125	− 97	− 34	− 121	+ 132
	113A	− 81	+ 159	− 94	− 33	− 137	+ 149
	124E	− 58	+ 161	− 68	− 36	− 169	+ 150
	141H	− 118	+ 118	− 115	− 32	− 115	+ 107
MLE	163H	− 147	+ 148	− 127	+ 2	− 127	+ 133
	220I	− 83	+ 132	− 73	− 34	− 159	+ 162
	267A	− 157	+ 156	− 111	+ 4	− 147	+ 130
PRIG-IGPS	81A	− 78	+ 157	− 93	− 34	− 148	+ 158
	131A	− 66	+ 149	− 109	− 1	− 164	+ 158
	178A	− 65	+ 135	− 88	− 27	− 128	+ 122
	230A	− 126	+ 161	− 90	− 9	− 173	+ 174
	276A	− 63	+ 147	− 90	− 48	− 111	+ 150
	328L	− 62	+ 145	− 84	− 28	− 155	+ 162
	373V	− 104	+ 119	− 82	− 26	− 157	+ 149
	421C	− 78	+ 169	− 67	− 29	− 176	+ 177
SUB	44V	− 96	+ 113	− 90	− 14	− 178	+ 185
	119M	− 67	+ 142	− 90	− 33	− 139	+ 130
	196L	− 72	+ 133	− 111	− 44	− 155	+ 162
Mean		− 87	+ 139	− 94	− 26	− 146	+ 147
Standard deviation		27	17	16	14	23	23
Conformation B							
CPA	59R	− 64	+ 145	− 44	+ 137	− 132	+ 160
	103M	− 132	+ 155	− 99	+ 155	− 126	+ 127
	238Y	− 103	+ 151	− 100	+ 155	− 142	+ 156
DFR	73A	− 109	+ 168	− 105	+ 110	− 98	+ 112
	92Q	− 109	+ 135	− 75	+ 144	− 99	+ 122
GBP	2D	− 68	+ 109	− 89	+ 124	− 102	+ 129
	85V	− 97	+ 135	− 70	+ 148	− 133	+ 126
	140Q	− 114	+ 148	− 89	+ 119	− 91	+ 109
	171I	− 107	+ 117	− 73	+ 142	− 131	+ 150
	230I	− 85	+ 127	− 77	+ 148	− 124	+ 135
GPD	71I	− 115	+ 131	− 81	+ 134	− 108	+ 126

(Continued)

Table 2 (Continued)

Protein	Initial Residue	ϕ_{-1}	ψ_{-1}	ϕ_1	ψ_1	ϕ_2	ψ_2
MLE	136L	− 135	+ 140	− 67	+ 137	− 128	+ 124
	193A	− 140	+ 153	− 92	+ 128	− 115	+ 125
	244A	− 101	+ 133	− 59	+ 156	− 101	+ 133
PRIG-IGPS	48T	− 62	+ 136	− 100	+ 155	− 126	+ 132
	109Q	− 70	+ 171	− 68	+ 143	− 109	+ 124
	158M	− 96	+ 148	− 86	+ 167	− 127	+ 137
	209V	− 103	+ 135	− 85	+ 122	− 104	+ 129
	305L	− 78	+ 178	− 101	+ 102	− 87	+ 141
	353V	− 81	+ 122	− 79	+ 162	− 117	+ 125
SUB	88V	− 62	+ 158	− 89	+ 118	− 106	+ 129
	147V	− 98	+ 131	− 83	+ 117	− 111	+ 121
	174V	− 116	+ 132	− 77	+ 115	− 89	+ 117
Mean		− 98	+ 142	− 82	+ 136	− 112	+ 129
Standard deviation		23	17	15	18	16	11
Other conformations							
CPA	199S	+ 153	0	+ 70	+ 60	− 127	+ 147
DFR	35V	− 56	+ 140	+ 76	− 6	− 111	+ 178
	55P	− 85	+ 166	+ 70	+ 22	− 157	+ 162
GBP	32P	− 55	− 32	− 80	− 8	− 102	+ 143
	103D	− 59	− 30	− 103	+ 25	− 104	+ 126
GPD	23	− 69	− 43	− 75	− 22	− 128	+ 163
MLE	295G	− 57	− 75	+ 121	+ 125	− 123	− 156
	323T	− 58	− 35	− 109	+ 7	+ 78	− 149
PRIG-IGPS	255G	+ 104	+ 171	− 87	+ 72	− 87	+ 73
	400G	− 64	− 14	− 127	+ 28	− 117	+ 134
SUB	25N	+ 74	+ 9	− 103	+ 135	− 102	+ 117
	264G	+ 64	− 144	− 72	− 25	+ 140	− 165

*Full protein names are given in table 1.

Conclusion

The observation of preferred conformations of strands entering parallel β -sheets exemplifies a style of investigation of protein architecture that has been made possible by the original demonstration that polypeptide conformations fall into discrete low-energy regions. If individual residues have only a few preferred conformations, short sequences of residues must have sets of preferred conformations based ultimately on – although not entirely determined by – the combinations of the allowed states of the individual residues. The many well-determined protein structures now available make it possible to identify which combinations of residue conformations are preferred in which structural contexts.

Acknowledgements

We thank Dr. J. N. Jansonius and Dr. A. Goldman for permission to use the unpublished atomic coordinates of PRIG-IGPS and MLE. A. M. L. thanks the Kay Kendall Foundation for generous support.

References

1. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein. Chem.* **23**, 284–437.
2. Rees, D. C., Lewis, M. & Lipscomb (1983) *J. Mol. Biol.* **168**, 367–387.
3. Bolin, J. T., Filman, D. J., Mathews, D. A., Hamlin, R. C. & Kraut, J. (1982) *J. Biol. Chem.* **257**, 13 650–13 662.
4. Vyas, N. K., Vyas, M. N. & Quijcho, F. A. (1988) *Science* **242**, 1290–1295.
5. Skarzynski, T., Moody, P. C. E. & Wonacott, A. J. (1987) *J. Mol. Biol.* **193**, 171–187.
6. Goldman, A., Ollis, D. L. & Steitz, T. A. (1987) *J. Mol. Biol.* **194**, 143–153.
7. Priestle, J. P., Grutter, M. G., White, J. L., Vincent, M. G., Kania, M., Wilson, E., Jardetzky, T. S., Kirshner, K. & Jansonius, J. N. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5690–5694.
8. Bode, W., Papamokos, E. & Musil, D. (1987) *Eur. J. Biochem.* **166**, 673–692.
9. Chothia, C. (1983) *J. Mol. Biol.* **163**, 107–117.

Analysis of Short Loops Connecting Secondary Structural Elements in Proteins

N. Srinivasan, R. Sowdhamini, C. Ramakrishnan and P. Balaram*

Molecular Biophysics Unit, Indian Institute of Science,
Bangalore 560012, India

The recognition by Ramachandran *et al.* (1) that the range of polypeptide chain conformations is limited by steric constraints constitutes the first step in defining the limits of the problem of protein folding. The crystal structures of globular proteins reveal intricate arrays of secondary structure modules, separated by irregular elements of the polypeptide chain termed as loops, which are uniquely assembled to yield the native folded conformation (2, 3). While much early work in the area of protein structure analysis focused on the regular elements of the structure, it is becoming increasingly apparent that the less regular segments or loops merit detailed study (4–10). Protein engineering approaches which permit transfer of loops from one context to another or allow truncation of connecting elements (11–13), have provided a fresh impetus to the detailed analysis of loops linking secondary structural elements in proteins. In the present study, short loops (≤ 5 residues) connecting secondary structure modules are identified from a data set of 65, largely non-homologous, protein crystal structures. Four types of super-secondary structural motifs involving α -helices and β -strands *viz.* $\alpha\alpha$, $\beta\beta$, $\alpha\beta$ and $\beta\alpha$ are considered and the conformational and compositional properties of the connecting loops are examined. An analysis of the spatial orientation of the two linked secondary structural elements reveals a reasonably frequent occurrence of an 'L'-shaped motif involving two orthogonally disposed β -strands.

Methods

The list of proteins used in the data set (14–77) is given in table 1. All proteins have been refined at a resolution of ≤ 2.0 Å. α -helices and β -strands were identified using the criterion that *at least four consecutive* residues should have backbone torsion angles in the following ranges:

α -helix: $\phi = -120^\circ$ to -10° , $\psi = -120^\circ$ to 20°

β -strand: $\phi = -180^\circ$ to -60° , $\psi = 60^\circ$ to 180° .

Figure 1 illustrates the boxes which define the helical and strand regions of (ϕ, ψ) space together with the distribution of experimentally determined (ϕ, ψ) values in the two

*For correspondence

Table 1. List of proteins used in the analysis.

S. No.	PDB code	Protein	Resolution (in Å)	Reference
1	1ACX	Actinoxanthin	2.0	14
2	1BP2	Phospholipase A2	1.7	15
3	1CPV	Ca binding parvalbumin	1.85	16
4	1CRN	Crambin	1.5	17
5	1CSE	Subtilisin carlsberg	1.8	18
6	1CTF	Ribosomal protein	1.7	19
7	1ECA	Hemoglobin (erythrocruorin)	1.4	20
8	1FB4	Immunoglobulin FAB	1.9	21
9	1FDX	Ferredoxin	2.0	22
10	1FX1	Flavodoxin	2.0	23
11	1GCR	γ-Crystallin	1.6	24
12	1GD1	Holo-glyceraldehyde-3-phosphate dehydrogenase	1.8	25
13	1GP1	Glutathione peroxidase	2.0	26
14	1HIP	High potential Fe protein	2.0	27
15	1HMQ	Hemerythrin	2.0	28
16	1INS	Insulin	1.5	29
17	1MBD	Myoglobin	1.4	30
18	1MLT	Melittin	2.0	31
19	1PCY	Plastocyanin	1.6	32
20	1PPT	Avian pancreatic polypeptide	1.37	33
21	1REI	Bence–Jones immunoglobulin	2.0	34
22	1RNT	Ribonuclease T ₁	1.9	35
23	1SN3	Scorpion neurotoxin	1.8	36
24	ITON	Tonin	1.8	37
25	1TPP	Trypsin	1.4	38
26	1UBQ	Ubiquitin	1.8	39
27	2ACT	Actinidin	2.0	40
28	2ALP	α-lytic protease	1.7	41
29	2APP	Acid proteinase	1.8	42
30	2AZA	Azurin	1.8	43
31	2B5C	Cytochrome b5	2.0	44
32	2CAB	Carbonic anhydrase B	2.0	45
33	2CDV	Cytochrome c3	1.8	46
34	2CNA	Conconavalin A	2.0	47
35	2CPP	Cytochrome P450	1.63	48
36	2CTS	Citrate synthase	2.0	49
37	2CYP	Cytochrome c peroxidase	1.7	50
38	2EBX	Erabutoxin	1.4	51
39	2LHB	Lamprey hemoglobin	2.0	52
40	2LZM	T4 lysozyme	1.5	53
41	2MHR	Myohemerythrin	1.7	54
42	2OVO	Ovomucoid third domain	1.5	55
43	2PAB	Prealbumin	1.8	56
44	2PKA	Kallikrein A	2.0	57
45	2RHE	Bence–Jones protein	1.6	58
46	2SNS	Staphylococcal nuclease	1.5	59
47	2SOD	Cu, Zn superoxide dismutase	2.0	60
48	2WRP	trp repressor	1.65	61
49	351C	Cytochrome c551	1.6	62
50	3BCL	Bacteriochlorophyll-A protein	1.9	63
51	3C2C	Cytochrome c2	1.68	64

Table 1. (Continued)

S. No.	PDB code	Protein	Resolution (in Å)	Reference
52	3DFR	Dihydrofolate reductase (<i>L. casei</i>)	1.7	65
53	3GRS	Glutathione reductase	1.54	66
54	3RP2	Rat mast cell protease	1.9	67
55	3RXN	Rubredoxin	1.5	68
56	3TLN	Thermolysin	1.6	69
57	3WGA	Wheat germ agglutinin	1.8	70
58	4DFR	Dihydrofolate reductase (<i>E. coli</i>)	1.7	65
59	4HHB	Human deoxyhemoglobin	1.74	71
60	4LDH	Lactate dehydrogenase	2.0	72
61	4TNC	Troponin C	2.0	73
62	5CPA	Carboxypeptidase A	1.54	74
63	5PTI	Bovine pancreatic trypsin inhibitor	1.0	75
64	5RSA	Ribonuclease A	2.0	76
65	6LYZ	HEW lysozyme	2.0	77

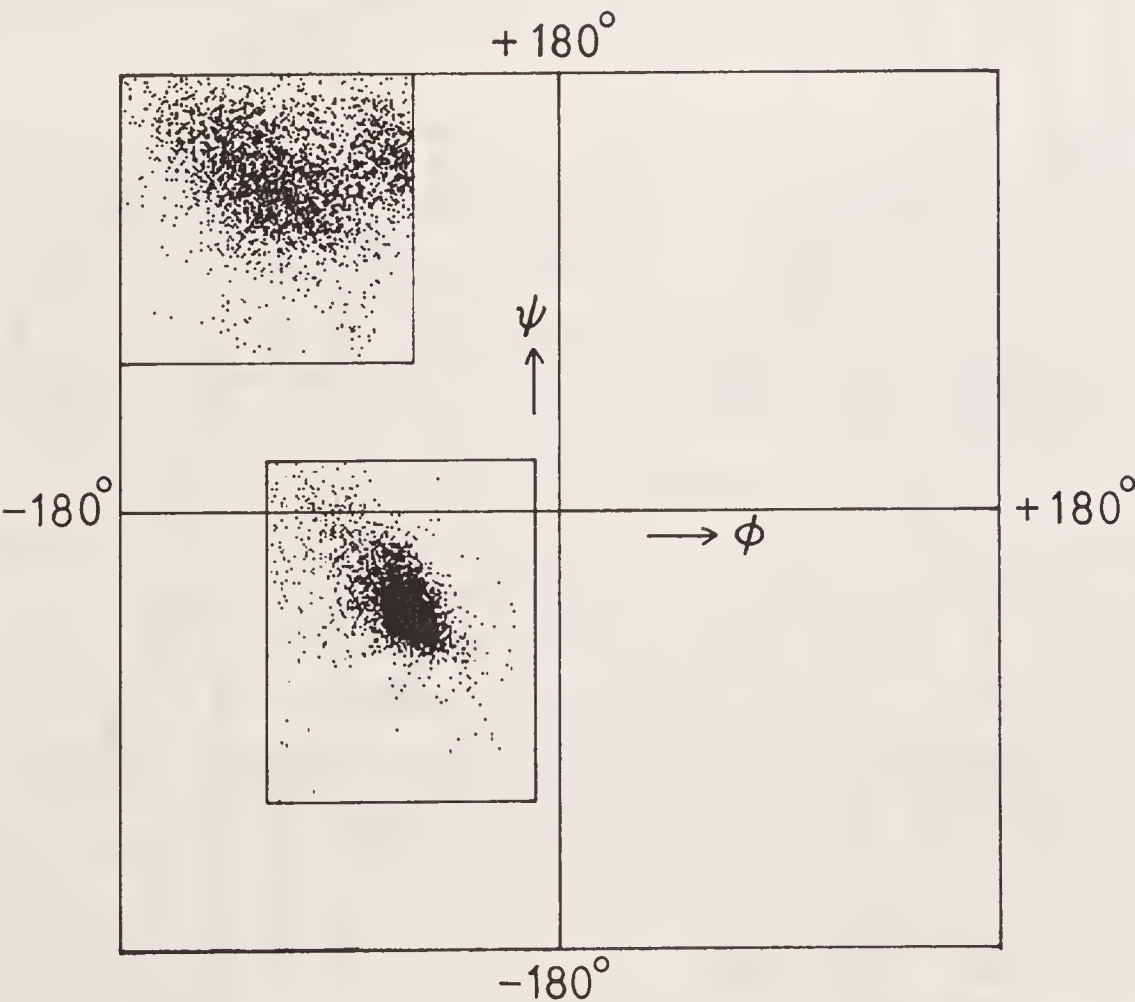


Figure 1. Boundaries of (ϕ, ψ) values used to identify helices and strands. The observed (ϕ, ψ) values in secondary structure elements identified in the 65 protein data set are also indicated. Note the clustering in the helical region and the greater dispersion for β -strands.

types of structures, identified in the data set. In the early stages of the analysis more rigorous limits were imposed on the dihedral angles, but were relaxed in later stages to accommodate moderate distortions. Similar limits of (ϕ, ψ) values were also used, for example, by Scheraga and coworkers (78) and Thornton and coworkers (8).

Super-secondary structure motifs were recognized when the connecting segment ranging between 1 and 5 residues did not conform to the above requirements. The number of motifs identified were: $\alpha\alpha$, 102; $\beta\beta$, 175; $\alpha\beta$, 77; $\beta\alpha$, 88. The maximum, minimum and mean number of residues in helices and β -strands present in these motifs are: α -helices, 32, 4 and 11; β -strands, 18, 4 and 6. Helix and strand axes were identified using a least squares fit procedure due to Chou *et al.* (79). The orientation of the two linked secondary structure elements is described by the angle between the corresponding axes. For an angle of 0° the two axes are antiparallel, while for 180° they are parallel.

Results and Discussion

Geometrical Parameters of Motifs

Figure 2 summarizes the distribution of the inter-secondary structure angle for the four motifs. For the $\alpha\alpha$ structures a large fraction of the observed examples have orientations between 30° and 90° , with only three examples of highly obtuse angles

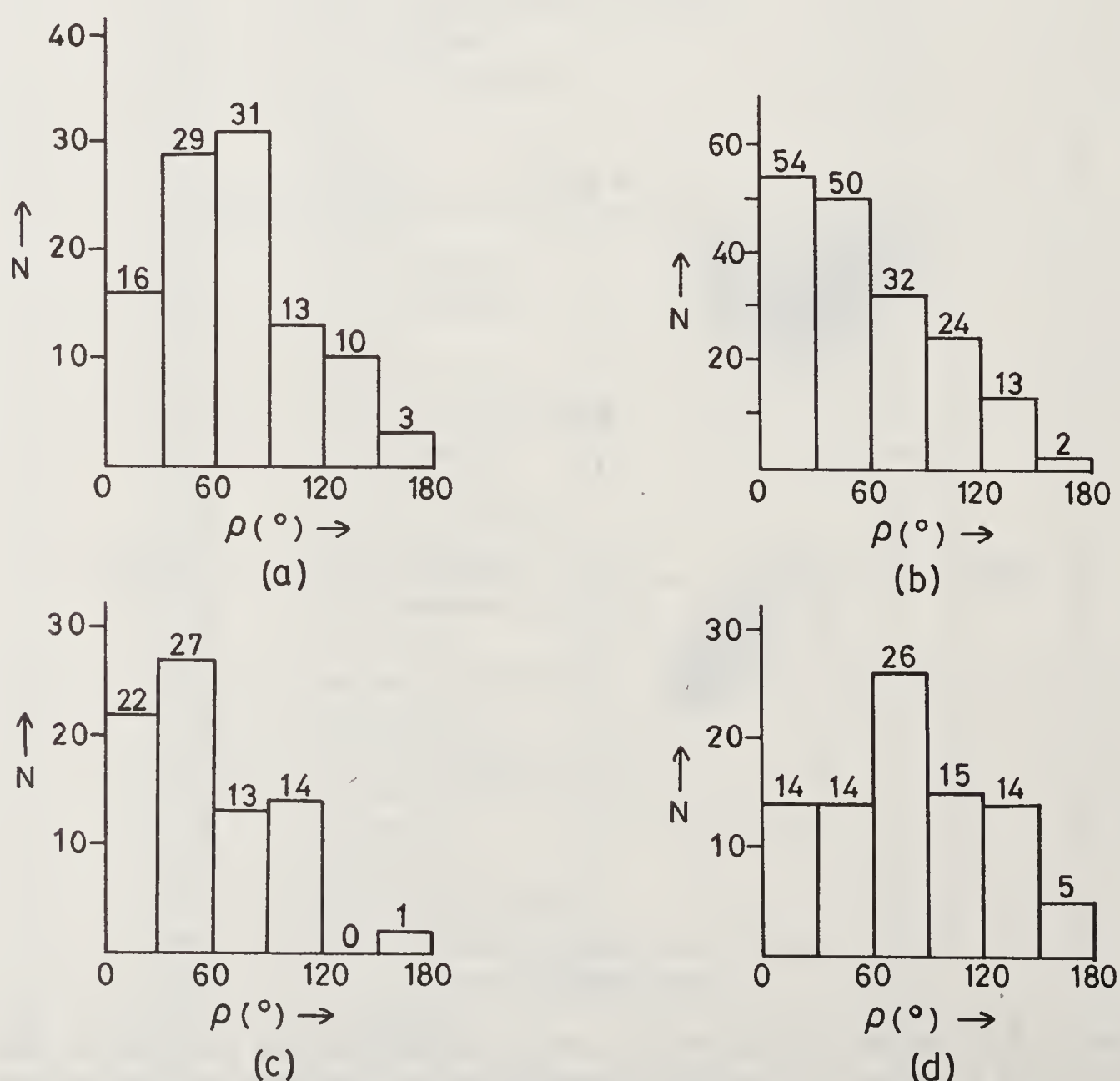


Figure 2. Distribution of the inter-module angle (ρ) for various super-secondary structural motifs in 65 proteins. (a) $\alpha\alpha$, (b) $\beta\beta$, (c) $\alpha\beta$, (d) $\beta\alpha$.

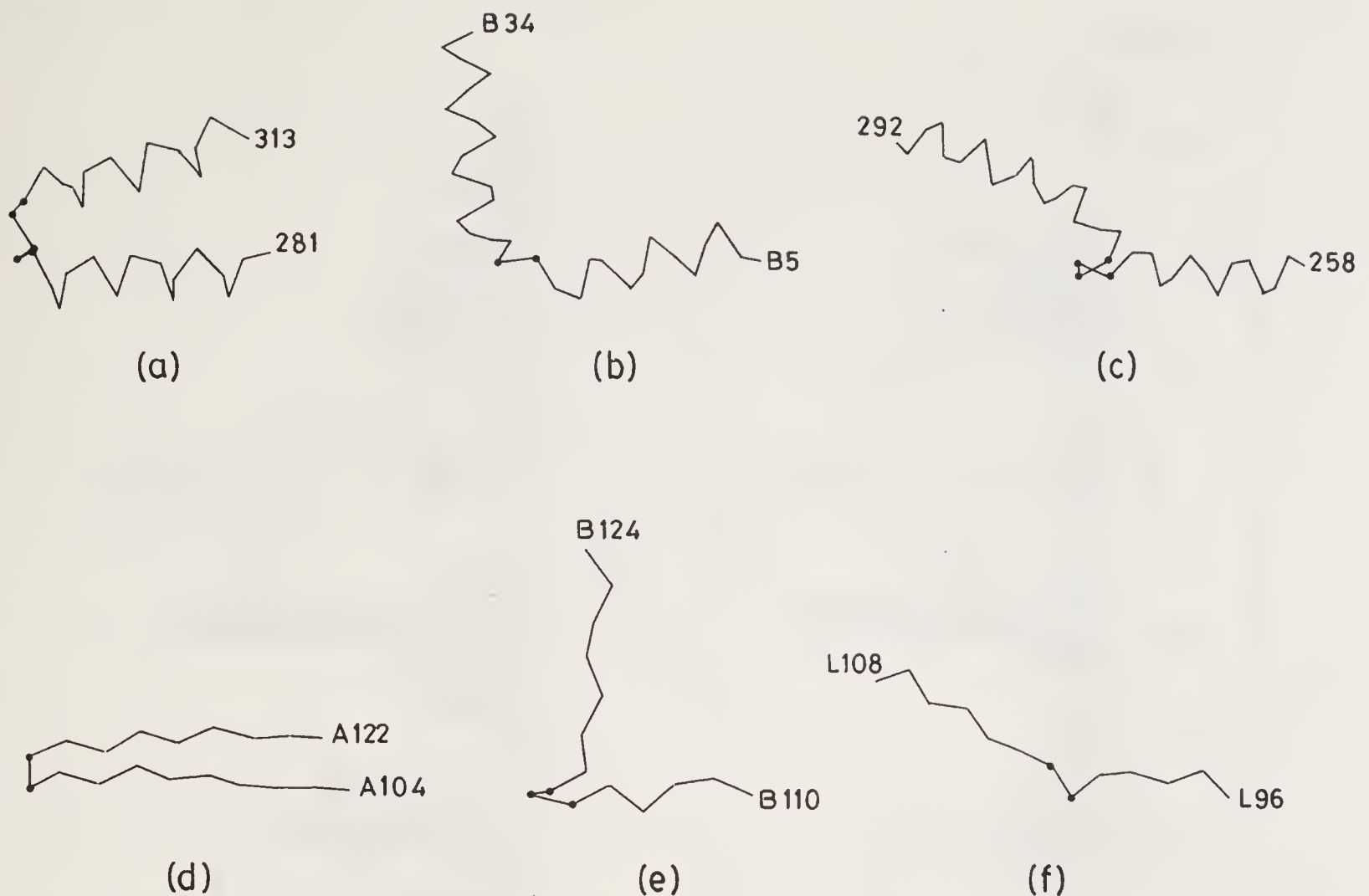


Figure 3. C^α plot of representative $\alpha\alpha$ and $\beta\beta$ motifs with distinct orientations of secondary structure modules. (a)–(c) $\alpha\alpha$ motifs; (d)–(f) $\beta\beta$ motifs. The first and last residue numbers defining the motif are indicated. Filled circles represent residue at the linker segment. (a) Thermolysin (3TLN) $\rho = 11.2^\circ$, (b) hemoglobin (4HHB) $\rho = 84.8^\circ$, (c) citrate synthase (2CTS) $\rho = 155.6^\circ$, (d) parvalbumin (2PAB) $\rho = 9.9^\circ$, (e) kallikrein (2PKA) $\rho = 86.5^\circ$, (f) immunoglobulin (1FB4) $\rho = 157.6^\circ$.

(> 150°). Thus, slightly kinked structures with two helices almost lying in a continuous line are rare. A large preponderance of roughly orthogonal helices is seen, a feature first described by Efimov (6), in an analysis of $\alpha\alpha$ corners. Figures 3a–c illustrate examples of the three distinct types of helix orientations seen for the short linker cases.

For $\beta\beta$ motifs, as expected, the preponderance of low values of the inter-strand angle (ρ), corresponds to the various antiparallel hairpin structures with short interconnecting loops. Interestingly, a significant number of $\beta\beta$ structures also adopt a right angled geometry leading to ‘L’ structures, which are considered later. Larger inter-strand angles corresponding to kinked structures are rare. Representative examples of the three major types of orientations are shown in figures 3d–f.

The distribution of inter-module angle for the $\alpha\beta$ and $\beta\alpha$ structures (figure 2c,d) reveal that the former tends to favour sharply turned structures, with only one example having $\rho > 120^\circ$. In contrast, $\beta\alpha$ motifs have a broader angular distribution with the peak observed between 60° – 90° corresponding to an approximately orthogonally oriented structure.

End-to-End Distance in Linking Loops

The distance between the last C^α atom of the amino terminus module and the first C^α atom of the carboxyl terminus module is defined as the end-to-end distance (E). For

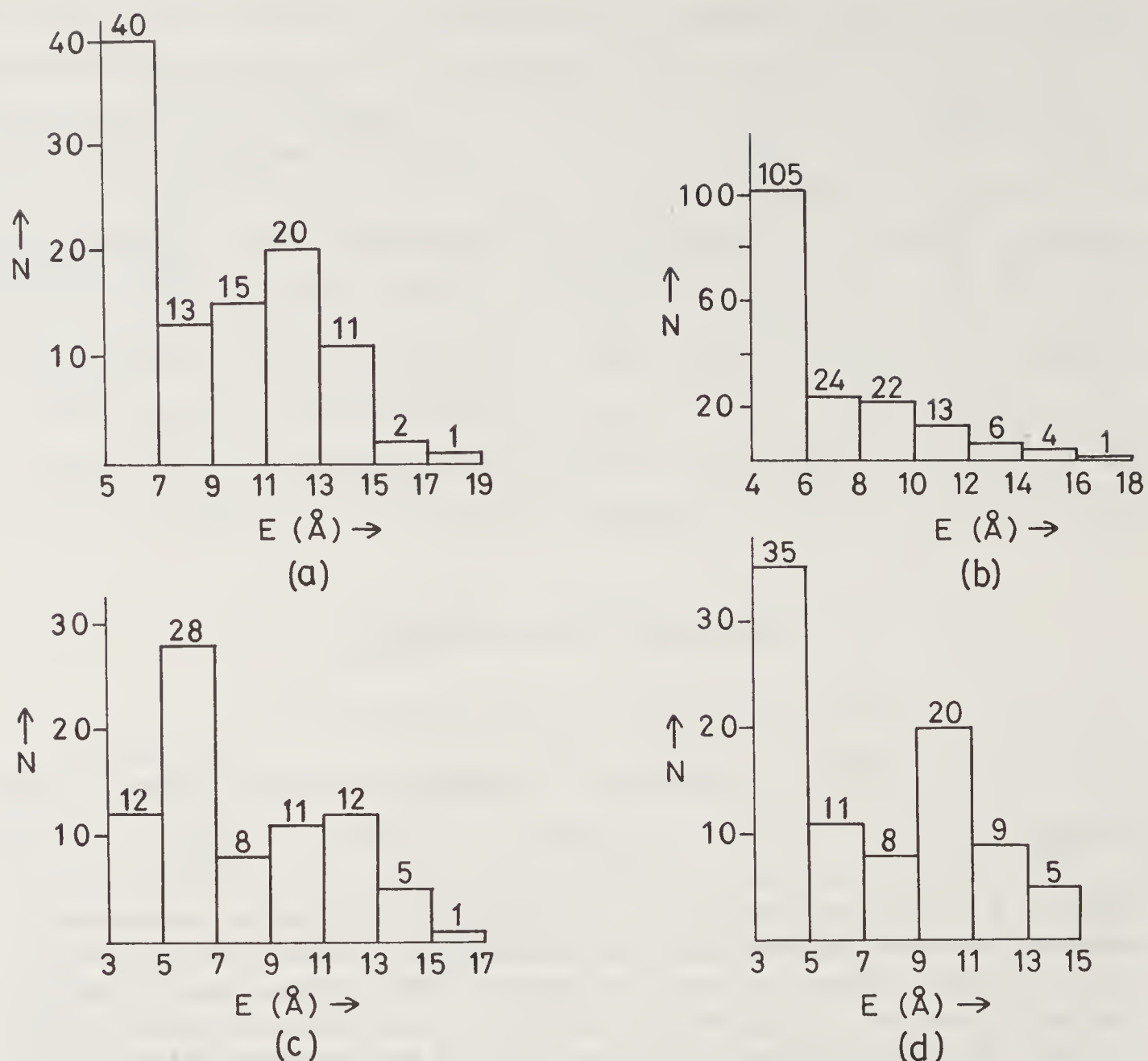


Figure 4. Distribution of end-to-end distance (E) of segments connecting secondary structural motifs in 65 proteins. (a) $\alpha\alpha$, (b) $\beta\beta$, (c) $\alpha\beta$, (d) $\beta\alpha$.

“zero-residue” linkers which are possible for $\alpha\beta$ and $\beta\alpha$ motifs, E should ideally be 3.8 \AA (the distance between C_i^α and C_{i+1}^α linker by a *trans* peptide unit). This is clearly borne out by the observations. For single-residue linkers E is a function of virtual bond angle $C_{i-1}^\alpha - C_i^\alpha - C_{i+1}^\alpha$ (δ), which has a restricted range of 74–145° (80), corresponding to sterically allowed conformations. E values in this case range between 4.6 and 7.2 \AA a feature again confirmed by observations.

For $\alpha\alpha$ structures, the most prominent peak in the distribution of E values (40 examples, figure 4a) occurs between 5 and 7 \AA , corresponding *exclusively* to single residue linkers. A second peak around 11–13 \AA is also seen. In the case of $\beta\beta$ motifs, a large majority (105 out of 175) examples have E between 4 and 6 \AA (figure 4b). 39 examples are single residue linkers, while 33 are two residue linkers. For $\alpha\beta$ motifs, all 12 examples between 3 and 5 \AA correspond to ‘zero-residue’ linkers. A peak is observed between 5 and 7 \AA (figure 4c). For $\beta\alpha$ motifs, all 35 examples below 5 \AA correspond to ‘zero-residue’ linkers. A significant peak between 9 and 11 \AA is also noted (figure 4d).

Amino Acid Propensities in Loops

The propensities of the various amino acids to occur in short loops between secondary structure modules was calculated in a manner similar to that used by Chou

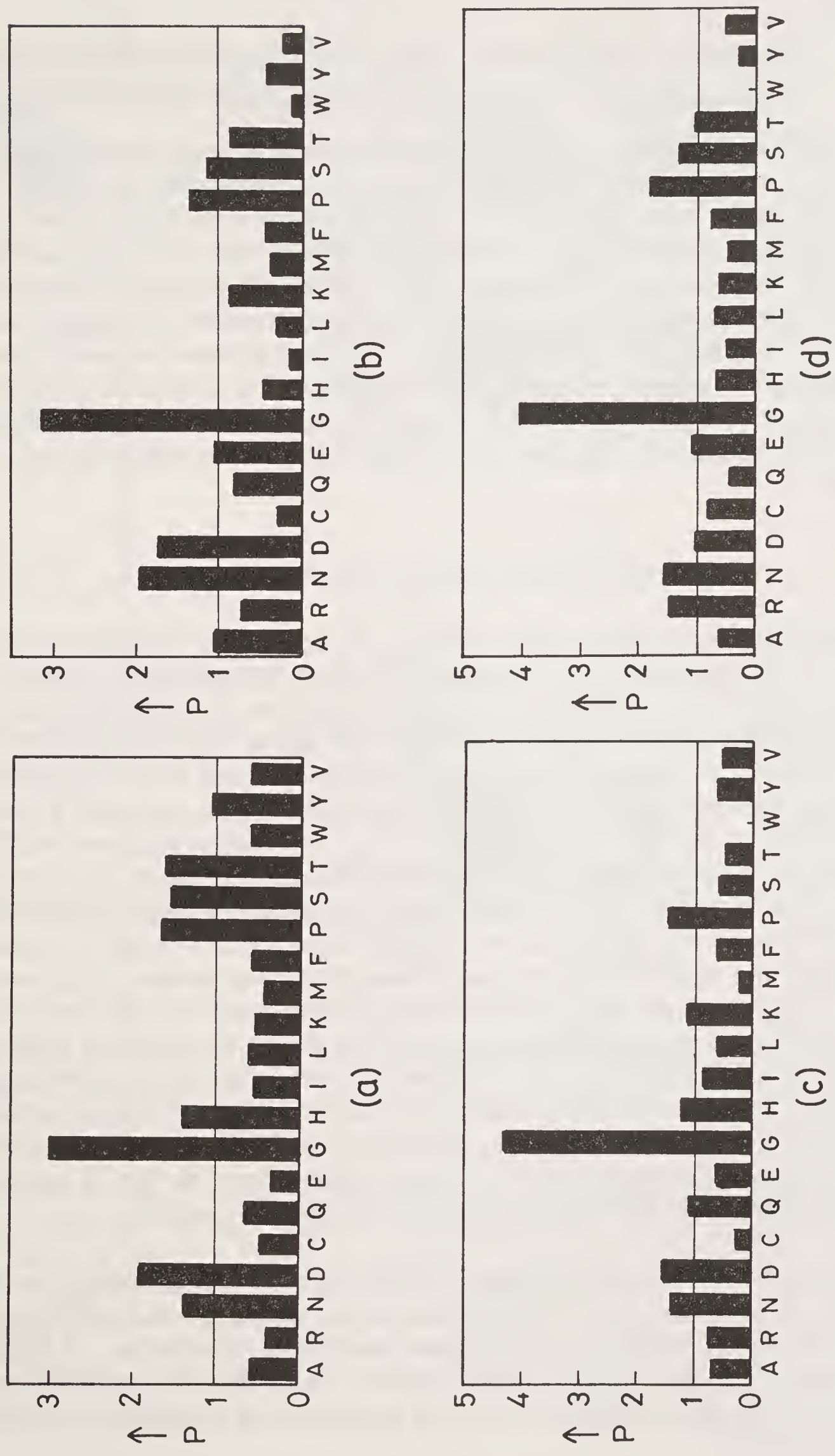


Figure 5. The propensity of occurrence of different amino acid residues in the linker segments of various motifs. (a) $\alpha\alpha$, (b) $\beta\beta$, (c) $\alpha\beta$, (d) $\beta\alpha$. Single letter codes of amino acids are indicated on the horizontal axis.

and Fasman for secondary structure preferences (81) and Leszczynski and Rose for loops (4):

$$\frac{\text{Propensity } P \text{ of amino acid Xxx in loop region}}{\text{Xxx in loop region}} = \frac{(\text{No. of Xxx in loops}/\text{No. of Xxx in motifs})}{(\text{No. of AA's in loops}/\text{No. of AA's in motifs})}$$

Residues with $P > 1.5$ have a very strong tendency to occur in loops, while those with $P < 0.5$ occur very infrequently in loops. Figures 5a–d summarize the P values for all twenty amino acids in the four types of connecting loops. It is clear that Gly is by far the most widely occurring loop residue. The small polar residues like Asp and Asn also show a significant tendency to occur in loops, particularly in $\alpha\alpha$ and $\beta\beta$ motifs. This is also true of Pro, which interrupts the sequential hydrogen bonding pattern in regular structures. Ser and Thr are found to a significant extent in the loop regions of $\alpha\alpha$ motifs but occur much less frequently in the other cases. Bulky hydrophobic residues like Trp, Ile, Leu and Val are also poorly represented in loops. The large, polar, charged residues Lys and Arg occur infrequently in loops. His is found to a significant extent in loops within $\alpha\alpha$ motifs.

Conformational Features of Linking Loops

Theoretical analysis of single residue linkers: In the simplest situation two linked secondary structure modules are connected by a single residue acting as a hinge. The orientation between the modules is then, only a function of the (ϕ, ψ) values at the linker. In order to examine whether local steric constraints, involving proximal atoms on the two modules, can limit the range of orientation (and linker conformation), computations were carried out on two linked 9 residue poly (Ala) segments. The linking residue was Gly or Ala and the conformation of the two 9-residue segments was fixed in ideal α -helical or β -strand regions, yielding $\alpha\alpha$ or $\beta\beta$ motifs. The results are summarized in figure 6. The variation of inter-module angle as a function of linker conformation is represented as a contour plot for the two motifs in figures 6a, d. Figures 6b, c show the sterically allowed region for the $\alpha\alpha$ case, where the linking residue is Ala and Gly, respectively. Similar results for the $\beta\beta$ motif are shown in figures 6e, f. It is clear that the range of conformational possibilities is significantly greater for the linker in $\beta\beta$ motifs as compared to $\alpha\alpha$ motifs. Indeed, in the former case the allowed region corresponds exactly to the Ramachandran steric map for an isolated residue. It may be noted that the observed (ϕ, ψ) values for single residue linkers (non-Gly) in the protein data set are somewhat more scattered than would be expected from figure 6b. This is presumably due to distortions from ideal helical conformation near the junction.

Conformational families in observed loops: Each of the four classes, namely, $\alpha\alpha$, $\beta\beta$, $\alpha\beta$ and $\beta\alpha$ linkers were sorted into groups, based on the number of residues. Analysis of (ϕ, ψ) values at the residues in the linkers resulted in recognition of consensus conformations of loops. Figures 7 and 8 show the (ϕ, ψ) distribution of the families recognized in this study and the C^α trace of representative examples, for $\alpha\alpha$ and $\beta\beta$ motifs, respectively.

$\alpha\alpha$ motifs: Four conformational families of $\alpha\alpha$ linkers were noticed. A strong family (23 out of 34) of single residue $\alpha\alpha$ linkers corresponds to extended or near extended

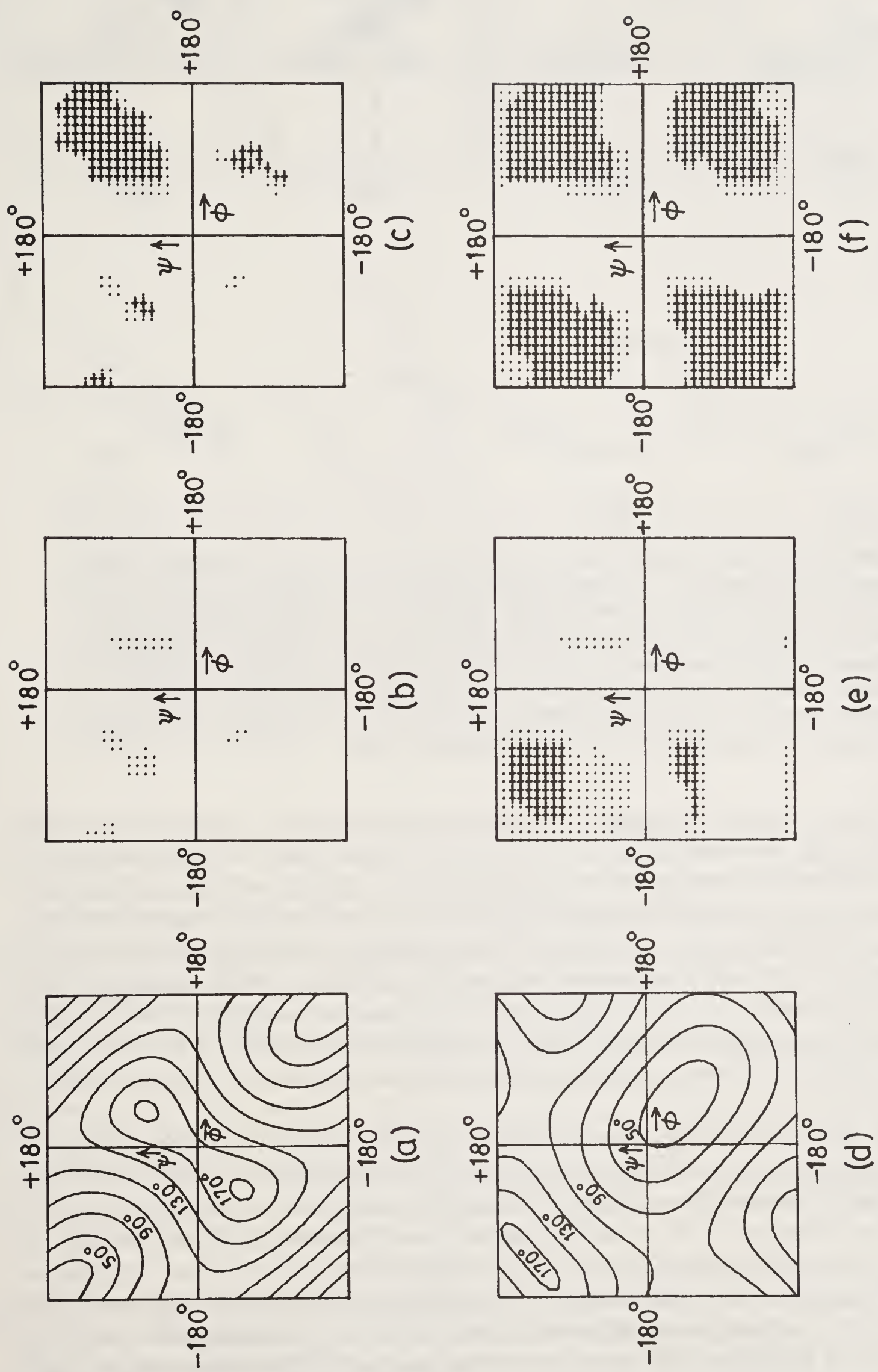


Figure 6. The orientation of secondary structure elements in $\alpha\alpha$ and $\beta\beta$ motifs connected by single residues, as a function of the (ϕ, ψ) values at the linker. Computations have been carried out for a system of two linked 9-residue poly (Ala) segments. Contour representations of the variation of the inter-element angle (ρ) are shown. (a) $\alpha\alpha$ motif; (b) $\beta\beta$ motif. The sterically allowed regions of (ϕ, ψ) space at the linker are mapped for Ala and Gly residues. (c) Ala, $\alpha\alpha$ (e) Gly, $\alpha\alpha$ (f) Gly, $\beta\beta$.

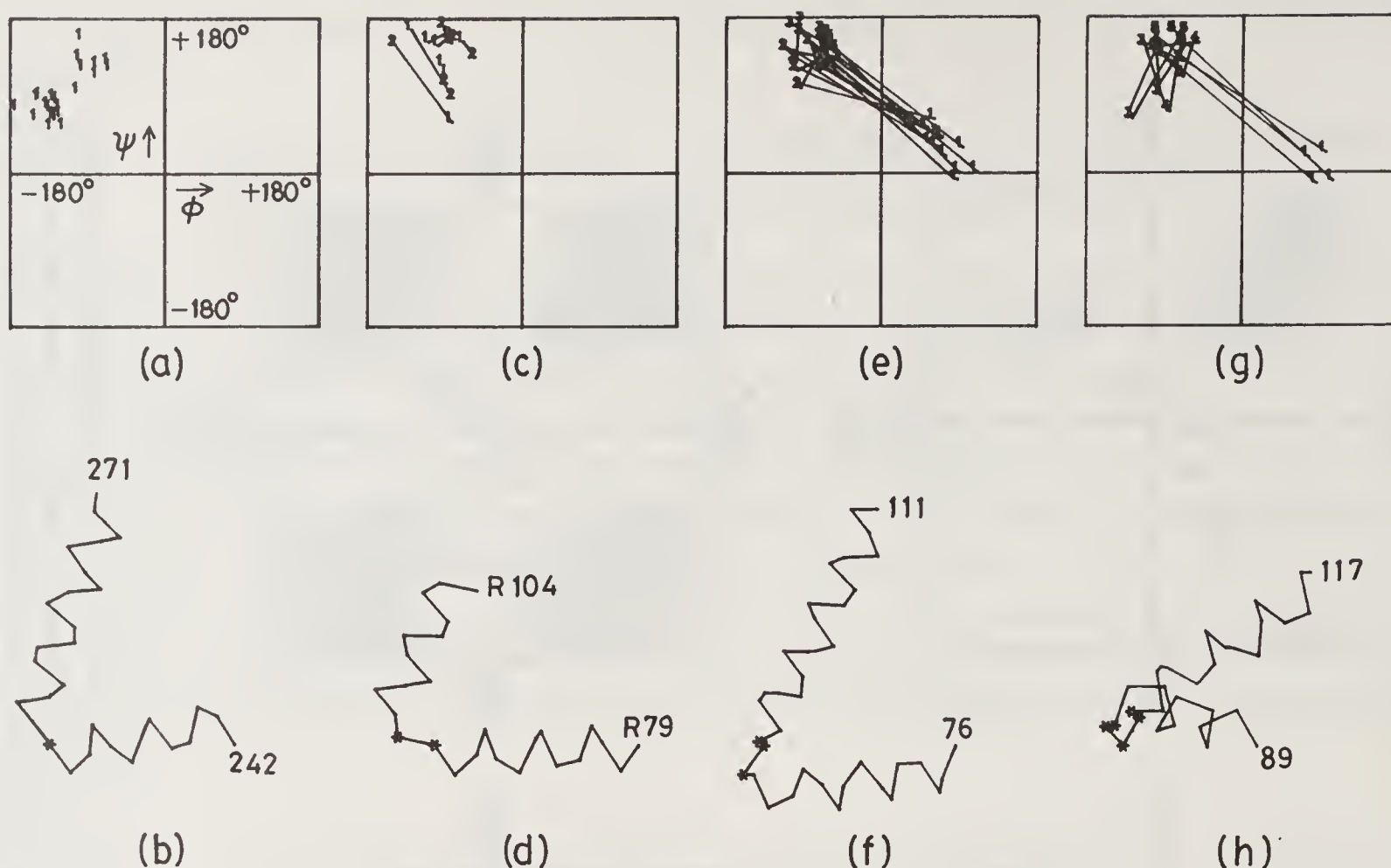


Figure 7. Conformational families observed for linking loops in $\alpha\alpha$ motifs, represented on a (ϕ, ψ) plot. C^α tracing of representative motifs with residue numbers indicated are shown below each plot. The numbers 1–5 in the (ϕ, ψ) diagrams mark the conformations at residues 1–5 of the linking segment. Residues in the linking loops are marked with an asterisk in the C^α tracings. (a) single residue linkers, (b) 242–271 cytochrome c peroxidase (2CYP), (c) two residue linkers, (d) R79–R104 *trp*-repressor (2WRP), (e) three residue linkers, (f) 76–111 erythrocrucorin (1ECA), (g) five residue linkers, (h) 89–117 citrate synthase (2CTS).

conformation at the linker (figure 7a). These examples would correspond to motifs where the angle between helices is nearly 60° (see figure 6a). The observations do support this feature. Seven of the 13 two residue $\alpha\alpha$ linkers have extended conformations at both the residues in the linker (figure 7c). A pronounced family in three residue linkers has the conformation $\alpha_L\beta\beta$, which is a feature in DNA binding helix-linker-helix motifs (see ref. 8). The α_L conformation is adopted by Gly in 6 out of 12 examples in this family. The 5 residue $\alpha\alpha$ linker family has a conformation $\alpha_L\beta\beta\beta\beta$. In all the four examples, Gly is the first residue (α_L) and the last residue in the linker is Thr. The three and five residue conformational families identified are observed in $\alpha\alpha$ corners (6).

$\beta\beta$ motifs: Two distinct conformational preferences are noticed for single residue $\beta\beta$ linkers, namely, α_R and α_L . The preference for α_R (24 examples) is more than that for α_L (10 examples). Several (23 examples) of these $\beta\beta$ motifs have an angle around 90° between the β -strands in the motif in agreement with the computed contour plot in figure 6d. In fact, there are 42 examples in the present data set, where two contiguous extended strands are disposed at about 90° to each other. Although a large number of such cases (25 examples) have single residue linkers, there are examples with more than one residue at the linker. The most interesting outcome of this analysis is the high frequency with which the linker residues take up (16 examples) type VIII β -turn conformations, first recognized by Wilmot and Thornton (82). Of these, 14 examples belong to $\beta\beta$ motifs with single residue linker with α_R conformation at the linker region.

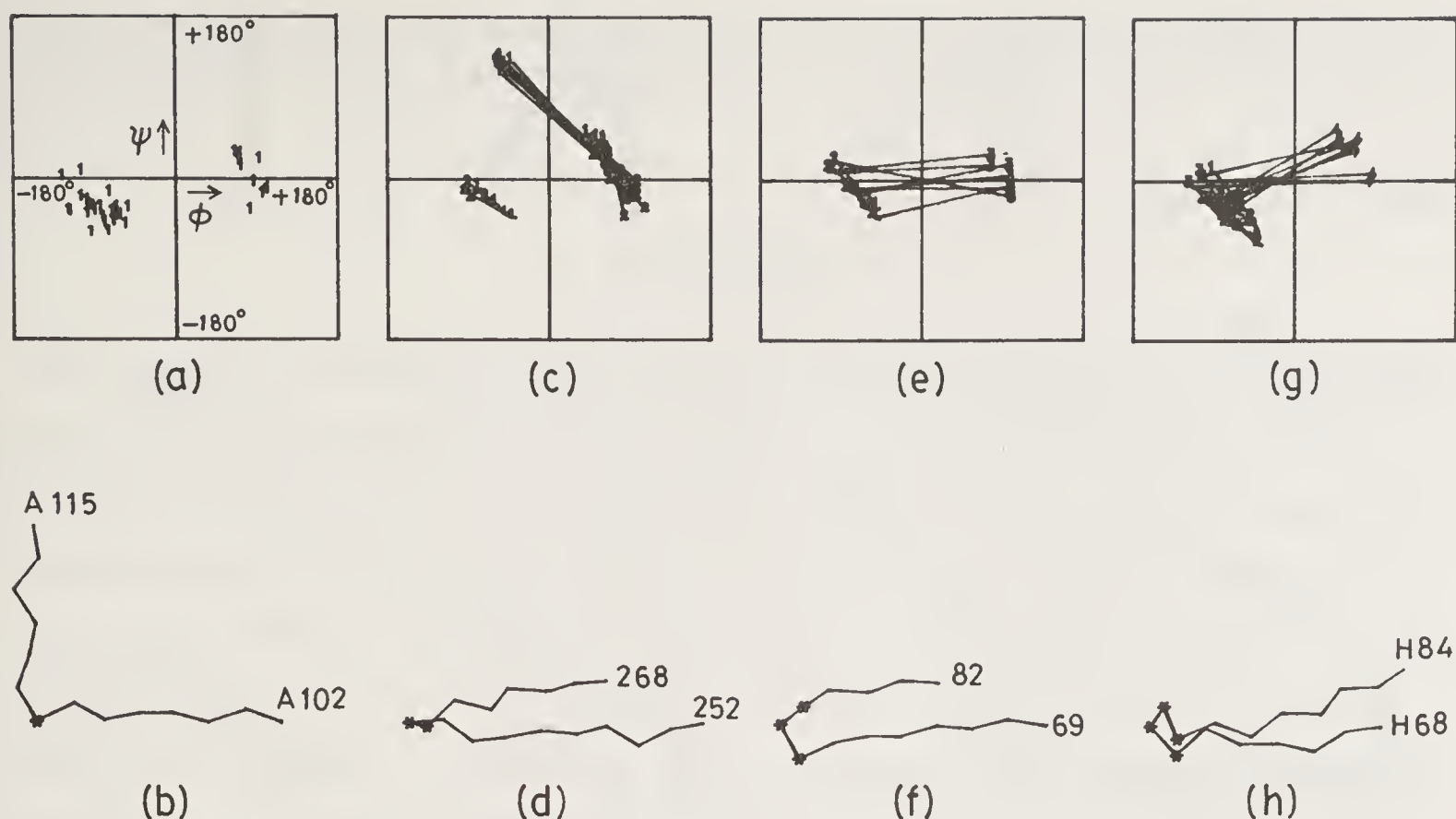


Figure 8. Conformational families observed for linking segments in $\beta\beta$ motifs represented as in figure 7. In case of more than one conformational family for a given number of residues in the linker, a representative C^α trace of the most populated family is shown. (a) Single residue linkers, (b) A102–A115 rat mast cell protease (3RP2), (c) two residue linkers, (d) 252–268 acid protease pencillopepsin (2APP), (e) three residue linkers, (f) 69–82 acid protease pencillopepsin (2APP), (g) four residue linkers, (h) H68–H84 immunoglobulin (1FB4). In the case of (c), the family $\beta\alpha_L$ is a result of marginal deviation of conformation at the first residue in the linker, from the one used to identify β -strands. If slightly relaxed criteria are employed, these examples would qualify as single residue $\beta\beta$ linkers.

More details on these L-shaped structures are presented elsewhere (R. Sowdhamini, N. Srinivasan, C. Ramakrishnan and P. Balaram, submitted for publication).

Two families are observed for two residue $\beta\beta$ linkers, *viz.*, $\alpha_L\alpha_L$ and $\alpha_R\alpha_R$. The $\alpha_L\alpha_L$ family is thickly populated (13 examples) and corresponds to a type I' β -turn at the linker resulting in a β -hairpin (5). Gly, Asn and Ser are generally present at these linker positions. The family $\alpha_R\alpha_R$ corresponds to type I β -turns. No residue preference is evident in this family. Three residue $\beta\beta$ linkers prefer $\alpha_R\alpha_R\alpha_L$ conformations. In 6 out of 7 members in this family the α_L conformation is adopted by Gly. The other two locations in this family are often taken up by Asp/Asn and Ser/Thr. Four residue $\beta\beta$ linkers prefer the conformation $\alpha_R\alpha_R\alpha_R\alpha_L$, with 8 examples in the data set. The α_L conformation is taken up by Gly, Asn or Asp. The α_R conformation in this family is frequently taken up by Ser or Asp or Asn. No specific conformational pattern is discernible for the five residue loops (23 examples).

$\alpha\beta$ and $\beta\alpha$ motifs: Figure 9 shows the (ϕ, ψ) distribution of $\alpha\beta$ and $\beta\alpha$ linker families. The C^α trace of representative examples for each are also shown.

Single residue $\alpha\beta$ linkers prefer α_L conformations. 15 out of 18 examples have Gly as the linker residue. A family of two residue $\alpha\beta$ linkers corresponds to $\beta\alpha_R$ conformations. No amino acid preference is evident in this family. Another such family corresponds to type II β -turns at the linker. Three out of four examples in this family have Gly at the

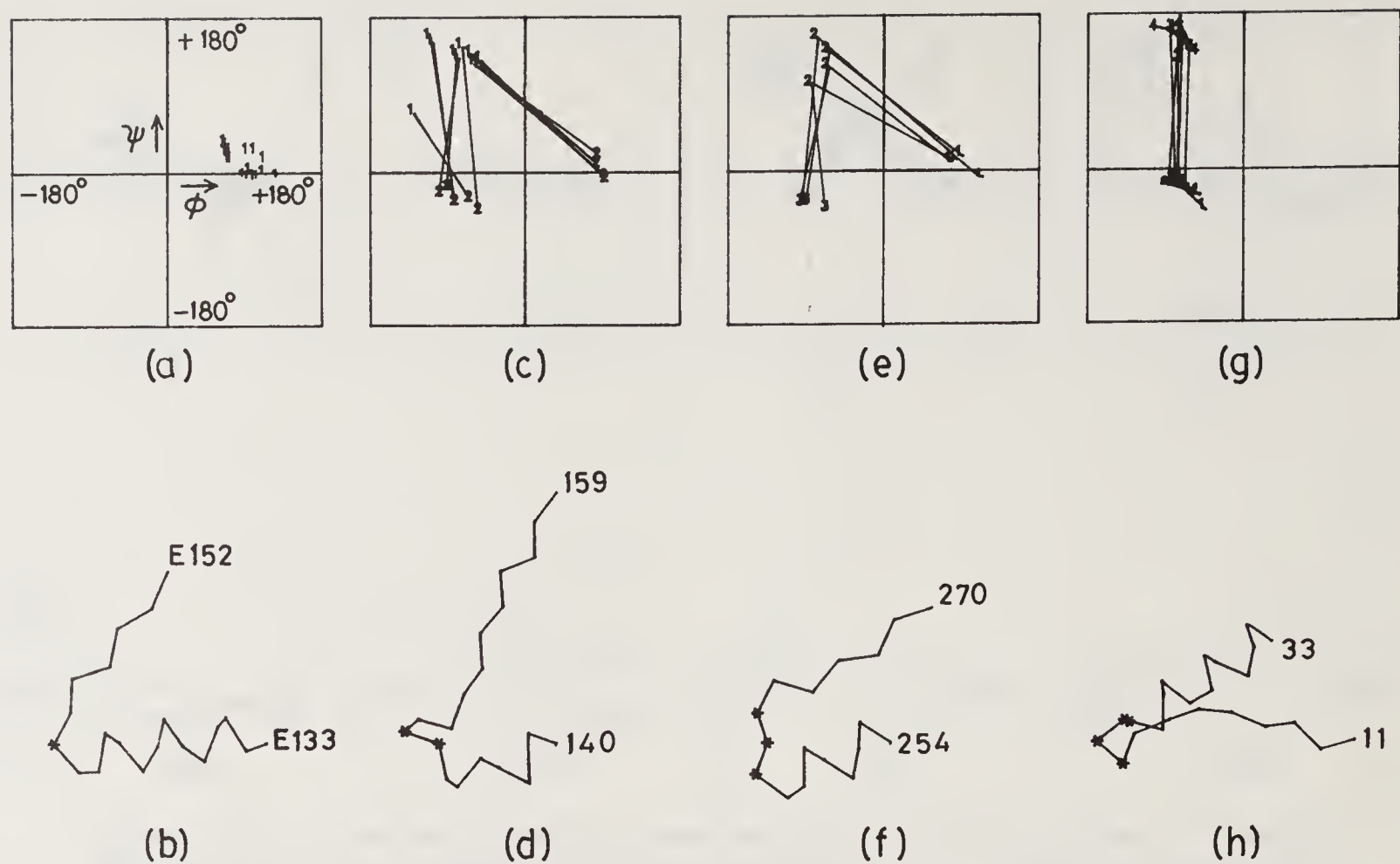


Figure 9. Conformational families observed for linking segments in $\alpha\beta$ (a–f) and $\beta\alpha$ (g, h) motifs represented as in figure 7. (a) Single residue linkers, (b) E133–E152 subtilisin carlsberg (1CSE), (c) two residue linkers, (d) 140–159 acid protease pencillopepsin (2APP), (e) three residue linkers, (f) 254–270 carboxypeptidase A (5CPA), (g) four residue linkers, (h) 11–33 ubiquitin (1UBQ).

second position of the linker. The three residue $\alpha\beta$ linker family corresponds to $\alpha_L\beta\alpha_R$ conformations. All the four examples in this family have Gly at the first position of the linker. Only one family is noticed for $\beta\alpha$ linkers and this corresponds to four residues in the linker region having α_R (α_R/γ_R) $\beta\beta$ conformations. No amino acid preference is noted in the present data.

Table 2. Conformational families of loops.

Loop type	No. of residues in the loop	Conformation at the loop	Number of examples
$\alpha\alpha$	1	β or near β	23
$\alpha\alpha$	2	$\beta\beta$	7
$\alpha\alpha$	3	$\alpha_L\beta\beta$	12
$\alpha\alpha$	5	$\alpha_L\beta\beta\beta\beta$	4
$\beta\beta$	1	α_R	24
$\beta\beta$	1	α_L	10
$\beta\beta$	2	$\alpha_R\alpha_R$	5
$\beta\beta$	2	$\alpha_L\alpha_L$	13
$\beta\beta$	3	$\alpha_R\alpha_R\alpha_L$	7
$\beta\beta$	4	$\alpha_R\alpha_R\alpha_R\alpha_L$	8
$\alpha\beta$	1	α_L	18
$\alpha\beta$	2	$\beta\alpha_R$	6
$\alpha\beta$	2	$\beta\alpha_L$	4
$\alpha\beta$	3	$\alpha_L\beta\alpha_R$	4
$\beta\alpha$	4	$\alpha_R\alpha_R/\gamma_R\beta\beta$	5

Table 2 summarizes observations on various conformational families of $\alpha\alpha$, $\beta\beta$, $\alpha\beta$ and $\beta\alpha$ linkers. Some of these families were also identified by Thornton and coworkers (5, 7–9).

Conclusions

This analysis of supersecondary structural motifs involving short connecting loops reveals that strong compositional and conformational preferences are observed for short linker segments. The results suggest that control of linker conformation may be an important element in protein design. The choice of specific loop sequences may aid in orienting secondary structure modules. Conformational regularities are harder to detect as the length of the linking segment increases. The frequent observation of α_L conformations in linker segments suggests that *de novo* design strategies involving chemical synthesis might use D-residues and other achiral amino acid analogs in putative loop segments. The analysis of orientational preferences reveals a significant population of orthogonal $\alpha\alpha$ and $\beta\beta$ motifs. Several cases of the L-shaped $\beta\beta$ motif provide examples of a type VIII β -turn at the hinge.

Acknowledgements

This research was supported by a grant from the Department of Science and Technology. RS was the recipient of a senior research fellowship from the Council of Scientific and Industrial Research.

References

1. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963) *J. Mol. Biol.* **7**, 95–99.
2. Richardson, J. S. (1981) *Adv. Protein Chem.* **34**, 167–339.
3. Richardson, J. S. & Richardson, D. C. (1989) in *Prediction of Protein Structure and Principles of Protein Conformation*, ed. Fasman, G. D. (Plenum Press, New York) pp. 1–98.
4. Leszczynski, J. F. & Rose, G. D. (1986) *Science* **234**, 849–855.
5. Sibanda, B. L. & Thornton, J. M. (1985) *Nature (London)* **316**, 170–174.
6. Efimov, A. V. (1984) *FEBS Lett.* **166**, 33–38.
7. Edwards, M. S., Sternberg, M. J. E. & Thornton, J. M. (1987) *Protein Eng.* **1**, 173–181.
8. Thornton, J. M., Sibanda, B. L., Edwards, M. S. & Barlow, D. J. (1988) *BioEssays* **8**, 63–69.
9. Sibanda, B. L., Blundell, T. L. & Thornton, J. M. (1989) *J. Mol. Biol.* **206**, 759–777.
10. Tramontano, P., Chothia, C. & Lesk, A. M. (1989) *Proteins* **6**, 382–394.
11. Hynes, T. R., Kautz, R. A., Goodman, M. A., Gill, J. F. & Fox, R. O. (1989) *Nature (London)* **339**, 73–76.
12. Wallace, C. G. A. (1987) *J. Biol. Chem.* **262**, 16767–16770.
13. Festow, J. S., Cardillo, T. S. & Sherman, F. (1989) *Proteins* **6**, 372–381.
14. Pletnev, V. Z., Kuzin, A. P., Trakharov, S., Popovich, V. & Tsigannik, I. (1982) in *Chemistry of Peptides and Proteins*, eds. Voelter, W. E., Wuensch, E., Ovchinnikov, J. & Ivanov, I. (Walter de Gruyter, Berlin) **1**, 429–444.
15. Dijkstra, B. W., Kalk, K. H., Hol, W. G. J. & Drenth, J. (1981) *J. Mol. Biol.* **147**, 97–123.
16. Moews, P. C. & Kretsinger, R. H. (1975) *J. Mol. Biol.* **91**, 201–228.
17. Hendrickson, W. A. & Teeter, M. (1981) *Nature (London)*, **290**, 107–113.
18. Bode, W., Papamokos, E. & Musil, D. (1987) *Eur. J. Biochem.* **166**, 673–692.

19. Leijonmarck, M. & Liljas, A. (1987) *J. Mol. Biol.* **195**, 555–579.
20. Steigemann, W. & Weber, E. (1979) *J. Mol. Biol.* **127**, 309–338.
21. Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. (1980) *J. Mol. Biol.* **141**, 369–391.
22. Adman, E. T., Sieker, L. C. & Jensen, L. H. (1976) *J. Biol. Chem.* **251**, 3801–3806.
23. Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3857–3860.
24. Wistow, G., Turnall, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P. & Blundell, T. (1983) *J. Mol. Biol.* **170**, 175–202.
25. Skarzynski, T., Moody, P. C. E. & Wonacott, A. J. (1987) *J. Mol. Biol.* **193**, 171–187.
26. Epp, O., Ladenstein, R. & Wendel, A. (1983) *Eur. J. Biochem.* **133**, 51–69.
27. Carter, C. W. Jr., Kraut, J., Freer, S. T., Xuong, N.-H., Alden, R. A. & Bartsch, R. G. (1974) *J. Biol. Chem.* **249**, 4212–4225.
28. Stenkamp, R. F., Sieker, L. C. & Jensen, L. H. (1983) *Acta Crystallogr.* **B39**, 697–703.
29. Dodson, E. J., Dodson, G. G., Hodgkin, D. C. & Reynolds, C. D. (1979) *Can. J. Biochem.* **57**, 469–479.
30. Phillips, S. E. V. (1980) *J. Mol. Biol.* **142**, 531–554.
31. Terwilliger, T. C. & Eisenberg, D. (1982) *J. Biol. Chem.* **257**, 6010–6015.
32. Guss, J. M. & Freeman, H. C. (1983) *J. Mol. Biol.* **169**, 521–563.
33. Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. & Wu, C.-W. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4175–4179.
34. Epp, O., Lattman, E. E., Schiffer, M., Huber, R. & Palm, W. (1975) *Biochemistry* **14**, 4943–4952.
35. Arni, R., Heinemann, U., Maslowska, M., Tokuoka, R. & Saenger, W. (1987) *Acta Crystallogr.* **B43**, 548–554.
36. Almassay, R. J., Fontecilla-Camps, J. C., Suddath, F. L. & Bugg, C. E. (1983) *J. Mol. Biol.* **170**, 497–527.
37. Fujinaga, M. & James, M. N. G. (1987) *J. Mol. Biol.* **195**, 373–396.
38. Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983) *Acta Crystallogr.* **B39**, 480–490.
39. Vijay Kumar, S., Bugg, C. E. & Cook, W. J. (1987) *J. Mol. Biol.* **194**, 531–544.
40. Baker, E. N. & Dodson, E. J. (1980) *Acta Crystallogr.* **A36**, 559–565.
41. Fujinaga, M., Delbaere, L. T. J., Brayer, G. D. & James, M. N. G. (1985) *J. Mol. Biol.* **184**, 479–502.
42. James, M. N. G. & Sielecki, A. R. (1983) *J. Mol. Biol.* **163**, 299–362.
43. Baker, E. N. (1988) *J. Mol. Biol.* **203**, 1071–1095.
44. Matthews, F. S., Argos, P. & Levine, M. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 387–392.
45. Kannan, K. K., Ramanadham, M. & Jones, T. A. (1984) *Ann. N.Y. Acad. Sci.* **429**, 49–55.
46. Higuchi, Y., Kusunoki, M., Matsuura, Y., Yasuoka, N. & Kakudo, M. (1984) *J. Mol. Biol.* **172**, 109–139.
47. Reeke, Jr., Becker, J. W. & Edelman, G. M. (1975) *J. Biol. Chem.* **250**, 1525–1547.
48. Poulos, T. L., Finzel, B. C. & Howard, A. J. (1987) *J. Mol. Biol.* **195**, 687–700.
49. Remington, S., Wiegand, G. & Huber, R. (1982) *J. Mol. Biol.* **158**, 111–152.
50. Finzel, B. C., Poulos, T. L. & Kraut, J. (1984) *J. Biol. Chem.* **259**, 13027–13036.
51. Bourne, P. E., Sato, A., Corfield, S. & Low, B. W. (1985) *Eur. J. Biochem.* **153**, 521–527.
52. Honzatko, R. B., Hendrickson, W. A. & Love, W. E. (1985) *J. Mol. Biol.* **184**, 147–164.
53. Weaver, L. H. & Matthews, B. W. (1987) *J. Mol. Biol.* **193**, 189–199.
54. Sheriff, S., Hendrickson, W. A. & Smith, J. L. (1987) *J. Mol. Biol.* **197**, 273–296.
55. Bode, W., Epp, O., Huber, R., Laskowski, M. Jr. & Ardelt, W. (1985) *Eur. J. Biochem.* **147**, 387–395.
56. Blake, C. C. F., Geisow, M. J., Oatley, S. J., Rerat, B. & Rerat, C. (1978) *J. Mol. Biol.* **121**, 339–356.
57. Bode, W., Chen, Z., Bartles, K., Kutzbach, C., Schmidt-Kastner, G. & Bartnuik, H. (1983) *J. Mol. Biol.* **164**, 237–282.
58. Furey, W. Jr., Wang, B. C., Yoo, C. S. & Sax, M. (1983) *J. Mol. Biol.* **167**, 661–692.
59. Legg, M. J. (1977) Thesis, Texas Agricultural and Mining University, U.S.A.
60. Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S. & Richardson, D. C. (1982) *J. Mol. Biol.* **160**, 181–217.

61. Lawson, C. L., Zhang, R.-G., Schevitz, R. W., Otwinowski, Z., Joachimiak, A. & Sigler, P. B. (1988) *Proteins* **3**, 18–31.
62. Matsuura, Y., Takano, T. & Dickerson, R. E. (1982) *J. Mol. Biol.* **156**, 389–409.
63. Tronrud, D. E., Schmid, M. F. & Matthews, B. W. (1986) *J. Mol. Biol.* **188**, 443–454.
64. Bhatta, G. (1981) PhD Thesis, University of California, San Diego.
65. Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C. & Kraut, J. (1982) *J. Biol. Chem.* **257**, 13650–13662.
66. Karplus, P. A. & Schulz, G. E. (1987) *J. Mol. Biol.* **195**, 701–729.
67. Reynolds, R. A., Remington, S. J., Weaver, L. H., Fisher, R. G., Anderson, W. F., Ammon, H. L. & Matthews, B. W. (1985) *Acta. Crystallogr.* **B41**, 139–147.
68. Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1980) *J. Mol. Biol.* **138**, 615–633.
69. Holmes, M. A. & Matthews, B. W. (1982) *J. Mol. Biol.* **160**, 623–639.
70. Wright, C. S. (1987) *J. Mol. Biol.* **194**, 501–529.
71. Fermi, G., Perutz, M. F., Shaanan, B. & Fourme, R. (1984) *J. Mol. Biol.* **175**, 159–174.
72. White, J. L., Hackert, M. L., Buehner, M., Adams, M. J., Ford, G. C., Lentz, P. J. Jr., Smiley, I. E., Steindel, S. J. & Rossman, M. G. (1976) *J. Mol. Biol.* **102**, 759–779.
73. Satyshur, K. A., Rao, S. T., Pyzalska, D., Drendel, W., Greaser, M. & Sundaralingam, M. (1988) *J. Biol. Chem.* **263**, 1628–1647.
74. Rees, D. C., Lewis, M. & Lipscomb, W. N. (1983) *J. Mol. Biol.* **168**, 367–387.
75. Wlodawer, A., Walter, J., Huber, R. & Sjolín, L. (1984) *J. Mol. Biol.* **180**, 301–329.
76. Wlodawer, A., Borkakoti, N., Moss, D. S. & Howlin, B. (1986) *Acta Crystallogr.* **B42**, 372–387.
77. Diamond, R. (1974) *J. Mol. Biol.* **82**, 371–391.
78. Isogai, Y., Nemethy, G., Rackovsky, S., Leach, S. J. & Scheraga, H. A. (1980) *Biopolymers* **19**, 1183–1210.
79. Chou, K.-C., Nemethy, G. & Scheraga, H. A. (1984) *J. Am. Chem. Soc.* **106**, 3161–3170.
80. Soman, K. V. (1984) PhD Thesis, Indian Institute of Science, Bangalore.
81. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 211–245.
82. Wilmot, C. M. & Thornton, J. M. (1988) *J. Mol. Biol.* **203**, 221–232.

Structural Studies on the Mechanism of O₂/Peroxide Activation and Electron Transfer in Peroxidase and Cytochrome P-450*

Thomas L. Poulos

Center for Advanced Research in Biotechnology of the Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850, USA, and the Department of Chemistry/Biochemistry, University of Maryland, College Park, USA

Introduction

Heme proteins have proven excellent subjects for detailed biophysical and structure/function studies. This is in part due to the unique spectroscopic features of the heme group. Indeed, various sophisticated methods such as time-resolved and resonance Raman spectroscopies have been developed for biological problems using heme proteins. Of particular interest is the use of these proteins for in-depth studies on the relationship between structure and function. The heme most widely used by various organisms, iron protoporphyrin IX, is found in cytochromes, globins, and several enzymes. Each of these proteins exhibit very different functions which are controlled by the interaction between heme and protein. Therefore, a detailed comparative analysis between heme proteins can provide valuable insights into how structure controls function.

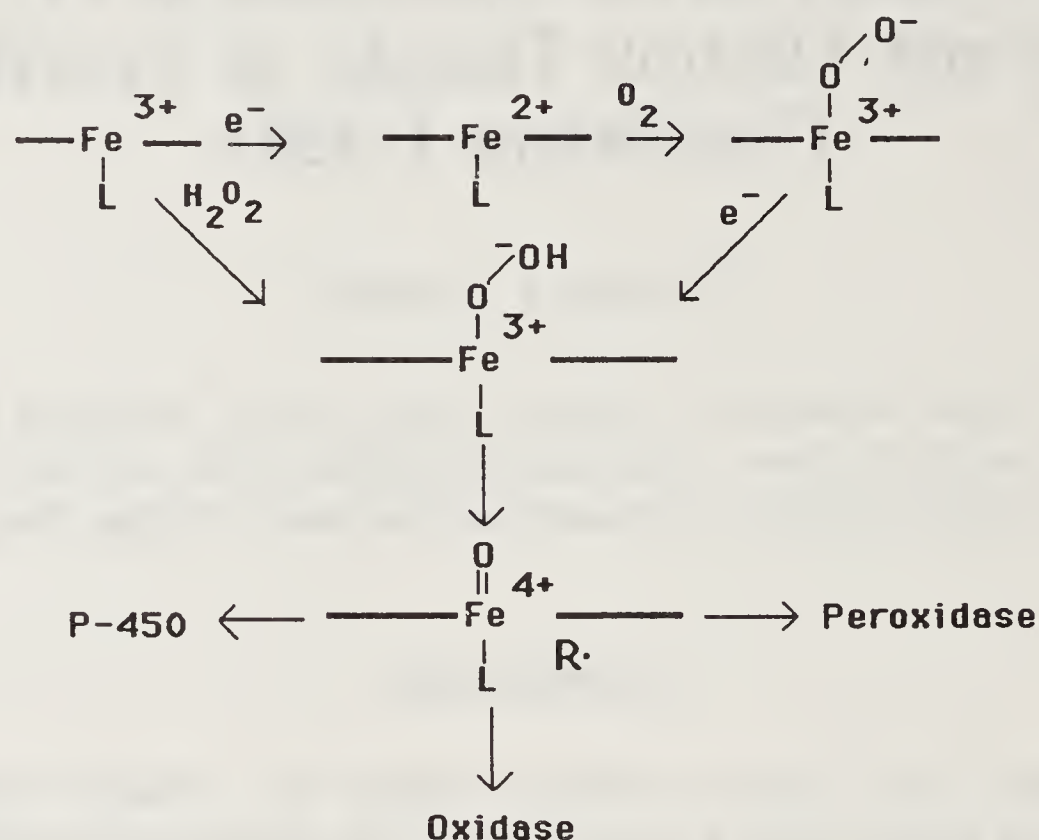
Although heme proteins are often considered as a single group, there actually are three distinct classes: oxygen carriers such as the globins, electron transport cytochromes, and heme enzymes. Heme enzymes catalyze the activation of oxygen and/or hydrogen peroxide in order to make the oxidizing equivalents of O₂ or H₂O₂ available for oxidation or hydroxylation of organic molecules and/or oxidation and electron transfer from another protein. Thus, the simplest way to look at a heme enzyme is in terms of two distinct reactions: first, activation of O₂ or H₂O₂ and second, electron transfer. It is often possible to study each step separately since in some cases the reaction product of the enzyme with O₂ or H₂O₂ is quite stable.

The focus of the present article is on what X-ray crystal structures have revealed about O₂/H₂O₂ activation and electron transfer reactions. Emphasis will be placed on two enzymes under active investigation in our laboratory: cytochrome c peroxidase (CCP) and cytochrome P-450.

*Work in the authors laboratory was supported by NSF Grant DMB 8716316 and NIH Grant GM 33688.

Generalized Heme Mechanism

The “generic” heme enzyme mechanism shown below provides a framework for understanding how these enzymes work.



Heme enzymes exhibit low (< 0 mV) redox potentials and thus prefer to remain in the Fe^{3+} oxidation state under “normal” physiological conditions. Since the charge on the porphyrin core is -2 and that of the iron $+3$, the net charge is $+1$. Control of the heme redox potential involves stabilization or destabilization of this extra positive charge by the axial ligand, L , and/or the immediate dielectric environment of the heme. The ligand, L , is an amino acid side chain. Crystallographic structures have shown that $L =$ histidine, tyrosine, or cysteine in peroxidase (1), catalase (2), and cytochrome P-450 (3, 4), respectively. The remaining axial coordination position is either vacant resulting in a high spin pentacoordinate system or is occupied by a dissociable solvent (water or hydroxide) molecule yielding hexacoordinate high or low spin complexes.

In oxidases and P-450 the first step in the reaction cycle is reduction of the ferric (Fe^{3+}) iron to ferrous (Fe^{2+}). This is necessary in order for molecular oxygen to bind. The electron transfer reaction often involves a protein electron donor thus necessitating formation of a pre-electron transfer protein–protein complex. That electron distribution shown for the oxy-complex in the above scheme is ferric iron and a superoxide anion which is supported by a series of observations (5–7). This reduction to the superoxide level is followed by a second reduction to the peroxide level. Presumably, protons accompany one or both electron transfer steps. In contrast to oxidases and P-450, peroxidases can go directly from the ferric resting state to the peroxy intermediate. Thus, all heme enzymes utilize a peroxy intermediate at some stage in their reaction cycle. The interesting chemistry begins with this peroxy intermediate. The O–O bond must be cleaved either heterolytically to release hydroxide (thereby leaving behind an oxygen atom) or homolytically to give two hydroxy radicals. The above scheme shows a heterolytic reaction since this is known to be the case for peroxidases (8); however, this is less clear for P-450 although a heterolytic mechanism is preferred by analogy with peroxidases and model heme studies. The

result is the storage of 2 oxidizing equivalents derived from O_2 or H_2O_2 in the enzyme active site so the formal oxidation state of the iron is +5. The above scheme shows +4 iron and a radical ($R\cdot$ in the above scheme), however, because in peroxidases one oxidizing equivalent resides on the heme macrocycle or an amino acid side chain. Where the second equivalent resides in the oxidases and P-450 is not known.

What happens next depends on the physiological role of the enzyme. In cytochrome oxidase, the oxidizing equivalents are used to oxidize cytochrome c in order to support mitochondrial electron transfer and oxidative phosphorylation. In P-450, the iron-linked oxygen atom of the activated intermediate is inserted into an organic substrate. Peroxidases also oxidize organic molecules, not by oxygen insertion but rather by an electron transfer mechanism involving a free radical. In contrast, cytochrome c peroxidase oxidizes ferrocyanochrome c. Catalase is similar to peroxidase except that a second H_2O_2 molecule is used as the reductant so the final product is molecular oxygen and water.

Oxygen and Peroxide Activation

Over the past decade, the high resolution refined crystal structures of cytochrome c peroxidase CCP (1, 9) catalase (2, 10) and cytochrome P-450 (4) have revealed some of the stereochemical details of heme enzyme mechanisms especially with respect to oxygen/peroxide activation (see 11 and 12 for previous reviews). The first heme enzyme crystal structure to be solved was CCP (1). At that time, cleavage of the peroxide O–O bond was known to involve acid/base catalyzed heterolysis. The reaction between H_2O_2 and the peroxidase proceeded to give distinct, semi-stable intermediates as shown below (8).



Compound I

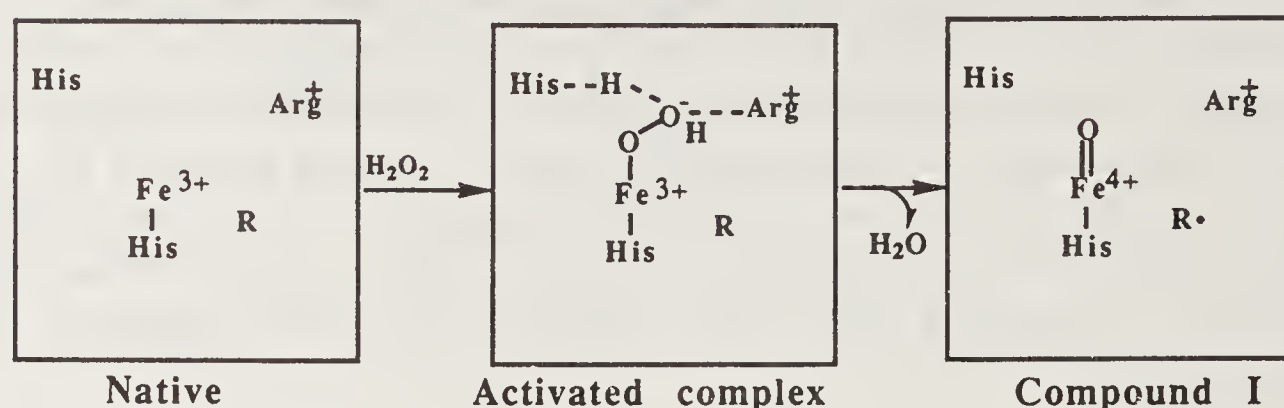


Compound II



The first step in the peroxidase reaction cycle is reaction between the enzyme and one mole of peroxide to give Compound I. Compound I is characterized by an oxy-ferryl iron atom (13, 14) and a free radical. Recently, the structure of CCP Compound I has been determined to 2.5 Å and has confirmed the presence of the ferryl-linked oxygen atom (15). In the plant peroxidases, the free radical is generally considered to be a porphyrin cationic radical (16) while in CCP it is clear that the radical site, R , is an amino acid side chain (17–19). Compound I is a powerful oxidizing agent and reacts with a substrate molecule in a one electron reaction to give Compound II and oxidized substrate. The porphyrin radical site in the enzyme is reduced in Compound I while aromatic substrates are oxidized to a free radical. Next, Compound II is reduced back to the native, ferric enzyme by a second substrate molecule. Catalase differs from peroxidases in that no distinct Compound II forms but instead a peroxide molecule reduces Compound I directly regenerating the resting ferric enzyme, and also producing molecular oxygen and water.

Many enzymological studies have focused on the mechanism of Compound I formation. The structure of CCP (figures 1 and 2) has revealed some of the stereochemical details of this reaction. Key to this mechanism is the structure of the protein surrounding the peroxide binding site on the distal side of the heme. Arg 48 and His 52 (CCP numbering) are conserved in the plant-like peroxidases and a catalytic role for these residues has been proposed as shown below.



In this scheme, the distal histidine operates as an acid–base catalyst by first abstracting a proton from the iron-linked peroxide oxygen atom and subsequently delivering it to the leaving oxygen atom. The arginine aids in stabilizing a negative charge on the leaving group that develops in the transition state. This catalytic machinery assures that the O–O bond cleaves heterolytically thereby releasing water while the enzyme retains a single iron-bound oxygen atom. The mechanism proposed for catalase Compound I formation based on the crystal structure is very similar (10). The main difference is that an asparagine residue plays the role of the distal arginine.

Immediately following cleavage of the O–O bond but preceding formation of a semi-stable Compound I, a single oxygen atom containing only six valence electrons remains linked to the iron. This hypothetical iron–oxygen intermediate is a powerful oxidizing agent. Extensive biophysical characterization of Compound I has shown that one electron is removed from the iron to give iron in the +4 oxidation state while a second electron is removed from *R* giving a radical. Thus, the iron-linked

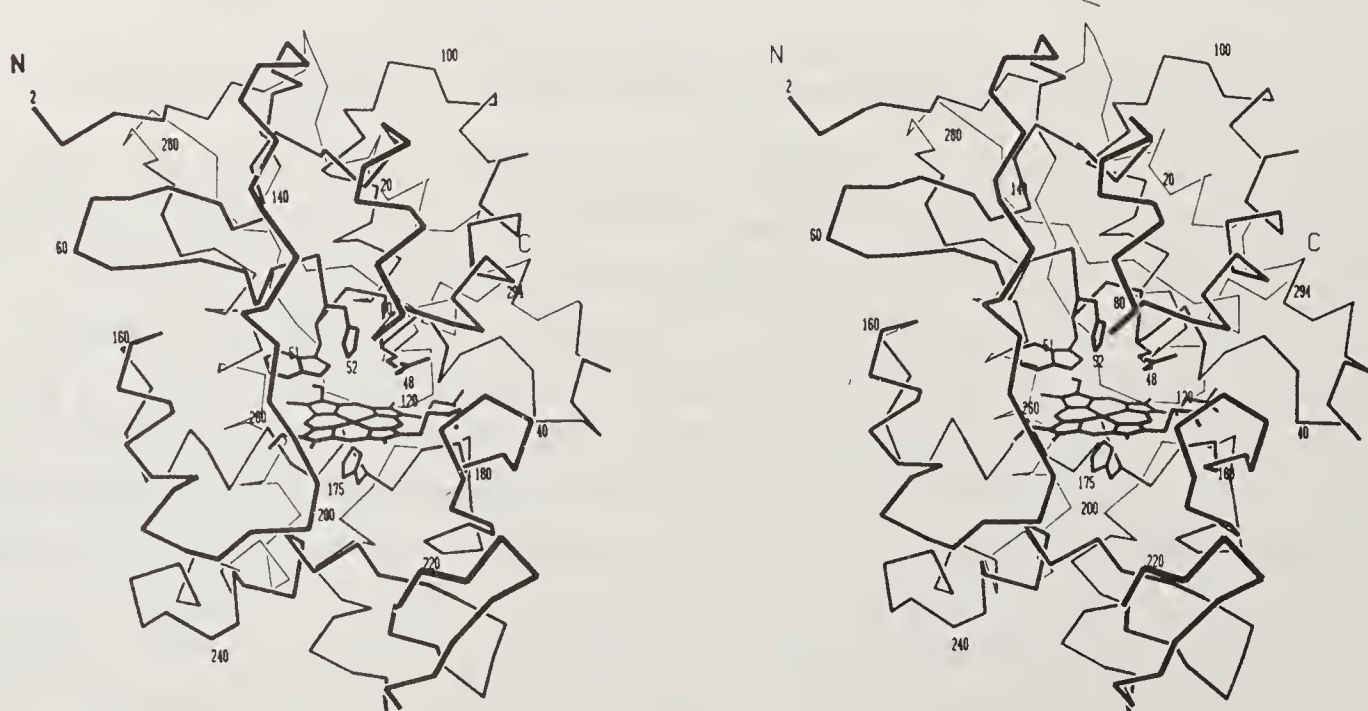


Figure 1. Stereoscopic view of the cytochrome c peroxidase molecule. The three distal residues, Arg 48, Trp 51, and His 52 and the proximal ligand, His 175, also are shown.

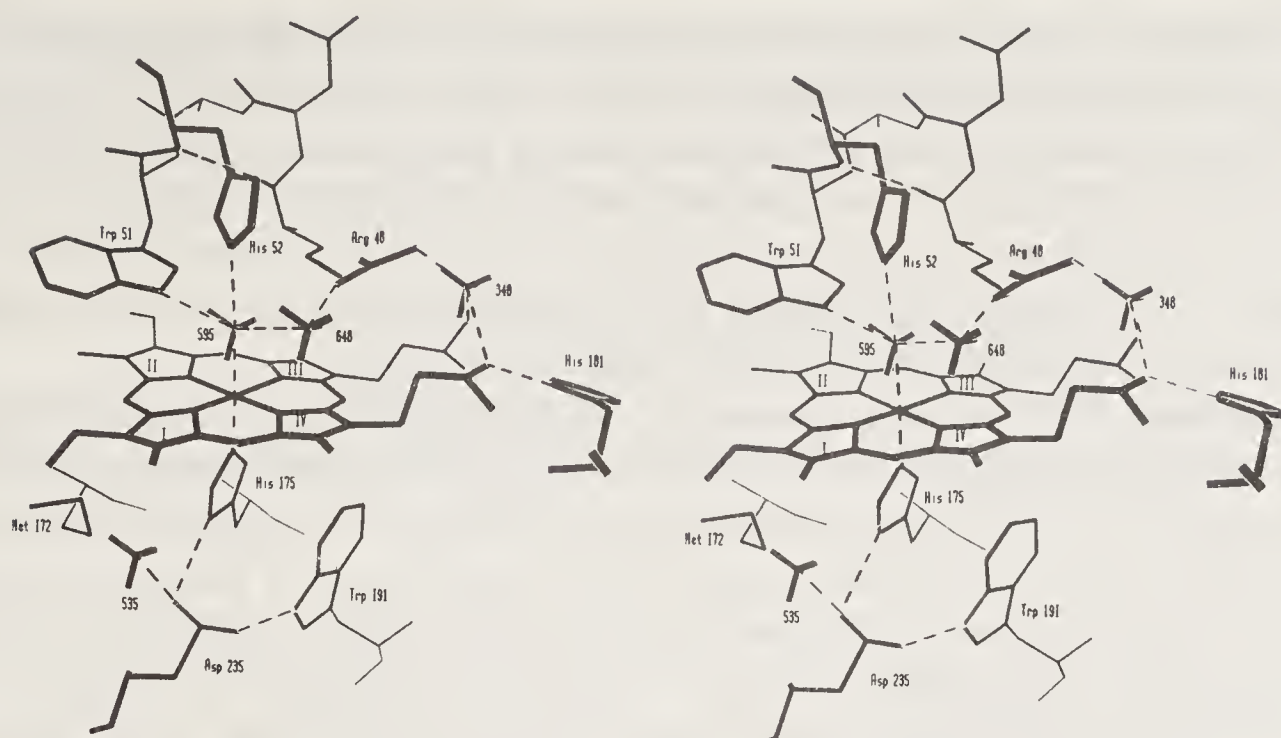
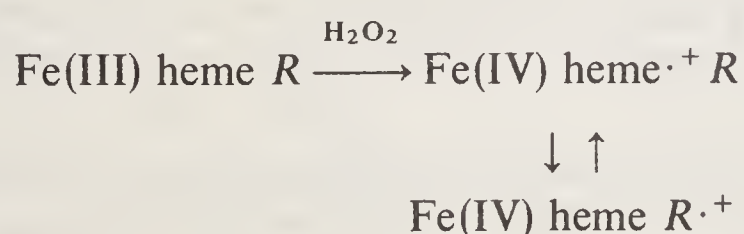


Figure 2. Stereoscopic view of the cytochrome c peroxidase active site. The tetrahedral objects indicate ordered solvent molecules and the dashed lines hydrogen bonds.

oxygen now contains a full complement of 8 valence electrons and is at the oxidation state of water. As stated earlier, in plant peroxidases R very likely is the heme itself while in CCP R is an amino acid side chain. How the precise location of oxidizing equivalents and charge distribution in Compound I is controlled by the interplay between the heme and its surrounding protein environment is one of the most interesting yet unsolved problems in peroxidase biophysics.

Mutagenesis and the Location of the CCP Radical

Since the discovery that CCP forms a relatively stable, amino acid centered radical, the identification of this residue or residues has been of central concern in understanding how CCP works. Early work in Yonetani's group (20) pointed toward aromatic residues, and most likely tryptophan. When the X-ray structure of CCP became available, Trp 51 was found to lie parallel to the heme directly adjacent to the peroxide binding site prompting us to propose that Trp 51 is the site of the radical (21, see figure 2). This notion was strengthened further by sequence alignments indicating that plant peroxidases contain a phenylalanine at this position rather than Trp. This led us to propose the following equilibria in the formation of Compound I (11).



In this scheme, the first product formed in CCP and other peroxidases upon reaction with peroxide is the cationic porphyrin radical followed by an equilibration between the porphyrin and an amino acid side chain, R . The relative stability of the side chain radical compared to the heme radical controls the predominant species in the above equilibrium. Because phenylalanine forms a much less stable radical than tryptophan,

the heme radical dominates in plant peroxidases while in CCP, the tryptophan radical is favored. However, the heme radical has never been observed in CCP presumably because the equilibration to the tryptophan radical was too fast to be measured even by stopped flow techniques. Nevertheless, a recent stopped flow study of a mutant CCP with diminished activity showed formation of the green, heme cationic radical characteristic of the plant peroxidases (22). Overall localization of the radical on Trp 51 seemed reasonable based on the available evidence.

Or so it seemed. Hoffman and coworkers using ENDOR spectroscopy concluded that a tryptophan could not be the radical site. Instead, they postulated a methionine-centered radical (23). The CCP structure supported this idea as well since the proximal pocket contains methionine residues close to the heme (figure 2). Moreover, magnetic circular dichroism spectra of CCP Compound I were interpreted as being incompatible with a tryptophan radical (24). Hori and Yonetani reexamined this question using single crystal EPR spectroscopy and found evidence once again for a tryptophan radical (25). The situation would have remained at somewhat of an impasse except for the addition of a new and powerful tool to the arsenal of techniques used to study CCP: site-directed mutagenesis.

The gene for CCP has been cloned (26, 27) and expressed in both yeast (28) and *E. coli* (29) allowing for the use of mutagenesis methods to test ideas about where the radical is located. Once the cloning, expression, and mutagenesis systems were established, some definitive answers were quickly forthcoming. Neither converting Trp 51 to phenylalanine nor removing a proximal methionine residue resulted in large changes in enzyme activity although the characteristic shape and stability of the Compound I EPR signal were altered (28, 29). However, converting Trp 191 to phenylalanine did dramatically decrease activity (30) and also significantly altered the stability and EPR spectral properties of Compound I. These results point to Trp 191 as a likely site for the free radical. Further support for a Trp 191 radical stems from ENDOR studies of CCP containing deuterated tryptophan residues (31). Incorporation of deuterated tryptophan caused a large change in the Compound I spectrum indicating that tryptophan is a critical component in the CCP radical. The single crystal EPR work of Hori and Yonetani (25) also is of interest here. They found that the axially symmetric g_{\parallel} component of the EPR signal of CCP Compound I is aligned along the heme normal and also found that the wavelength dependence of photochemical perturbation of the Compound I radical closely resembles what one expects for a tryptophan radical. Then if tryptophan is the radical site the indole ring must be either parallel or perpendicular to the heme plane since g_{\parallel} of the radical signal is aligned along the heme normal. Only Trp 51 and 191 have the indole rings either parallel or perpendicular to the heme normal. Since the mutagenesis work has demonstrated that Trp 51 is not essential, this leaves Trp 191 as the only other reasonable candidate.

Recent crystallographic studies further point to Trp 191 as an important site (32, 33). Based on 1.85 Å refined structures, it was found that nitric oxide (NO) coordinates with the CCP iron and specifically alters the Trp 191 region of CCP by causing an increase in crystallographic B or thermal factors and leads to a slight movement of Trp 191 away from the proximal ligand, His 175 (see figure 2). Other ligands like CN⁻, CO, and F⁻ do not have this effect. Of these ligands, NO is unique since it is paramagnetic and the unpaired spin is transferred to the heme system (34). We have postulated that NO imparts its effect on Trp 191 by delocalization of electron density

via the proximal ligand, His 175. Whatever the precise explanation of why NO perturbs Trp 191, it is clear that Trp 191 is uniquely sensitive to the electronic state of the heme which lends some support to conclusions derived from the mutagenesis work that Trp 191 plays a special role in catalysis.

The P-450 Oxygen Activation Reaction

After solution of the CCP and catalase structures, the P-450 cam structure became available (3,4 and figures 3 and 4). P-450 cam is purified from *Pseudomonas putida*

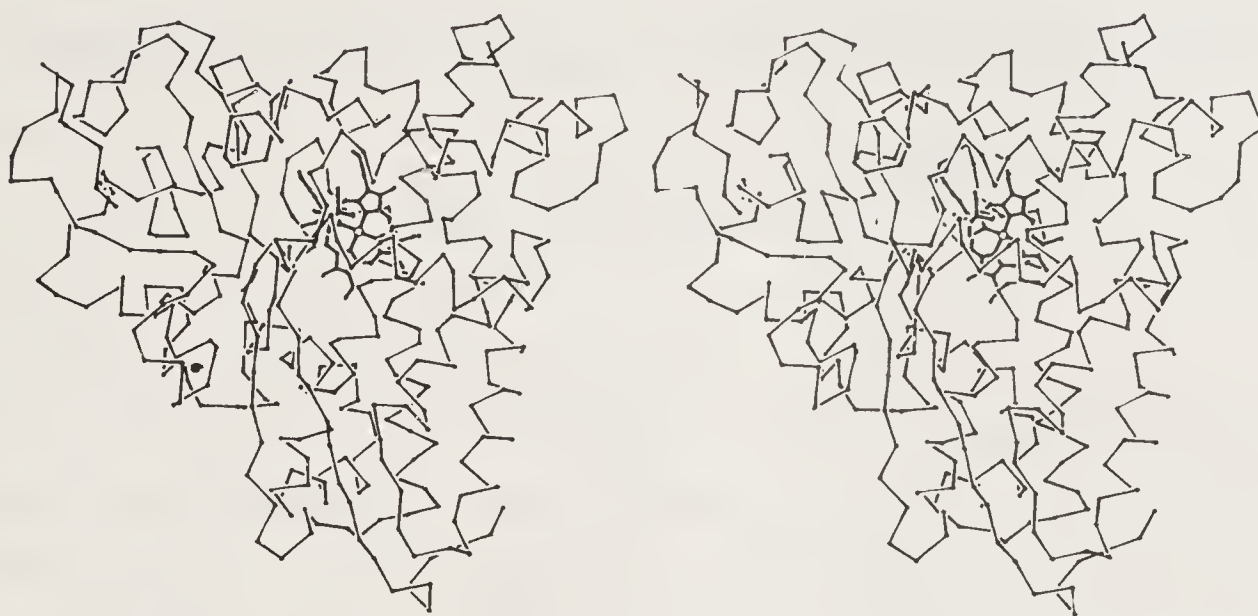


Figure 3. Stereoscopic view of the cytochrome P-450cam molecule showing the heme and camphor.

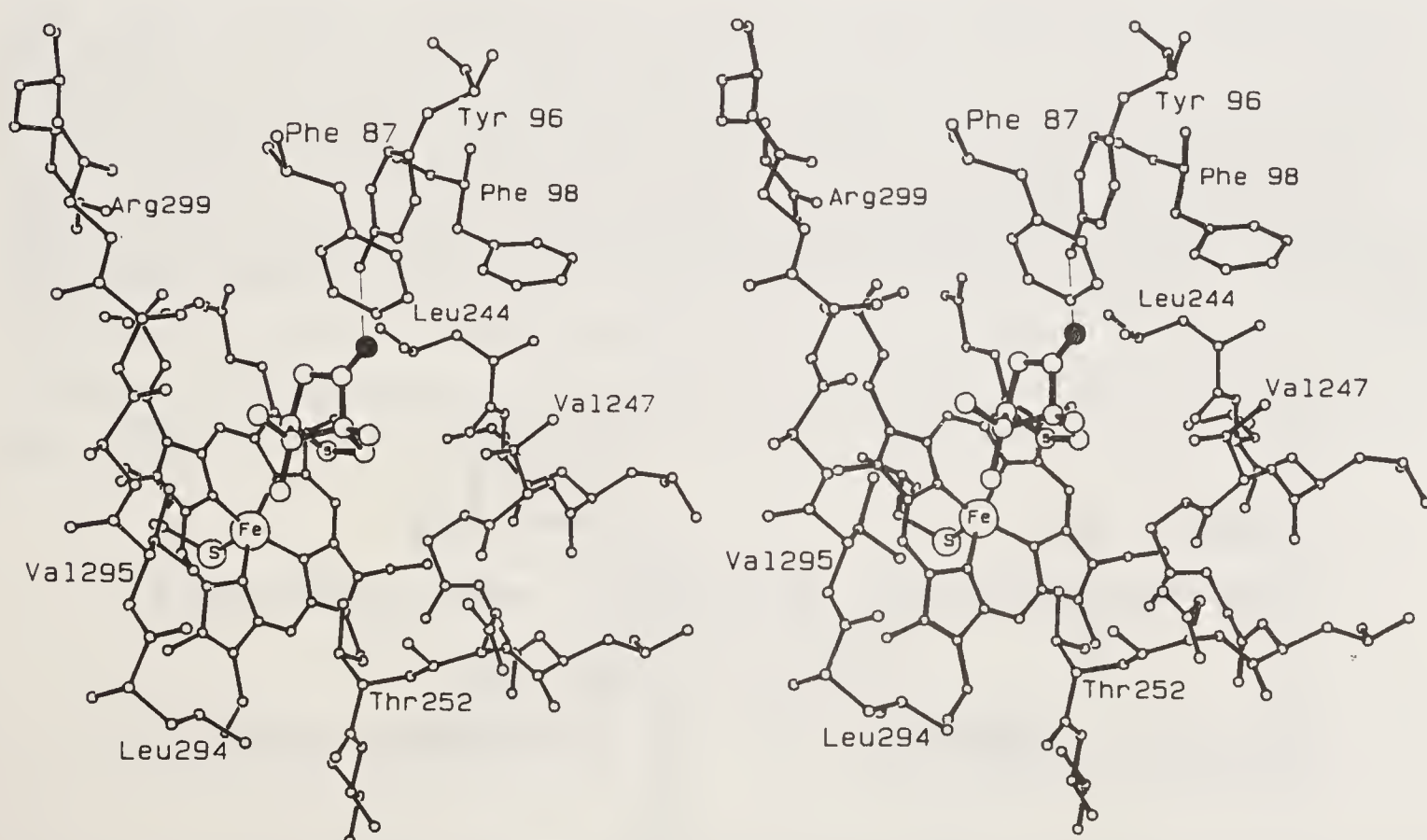
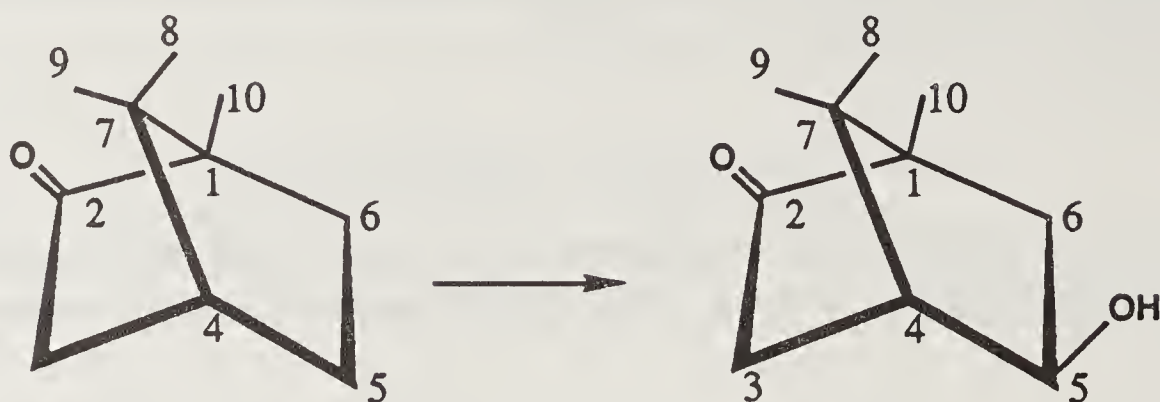
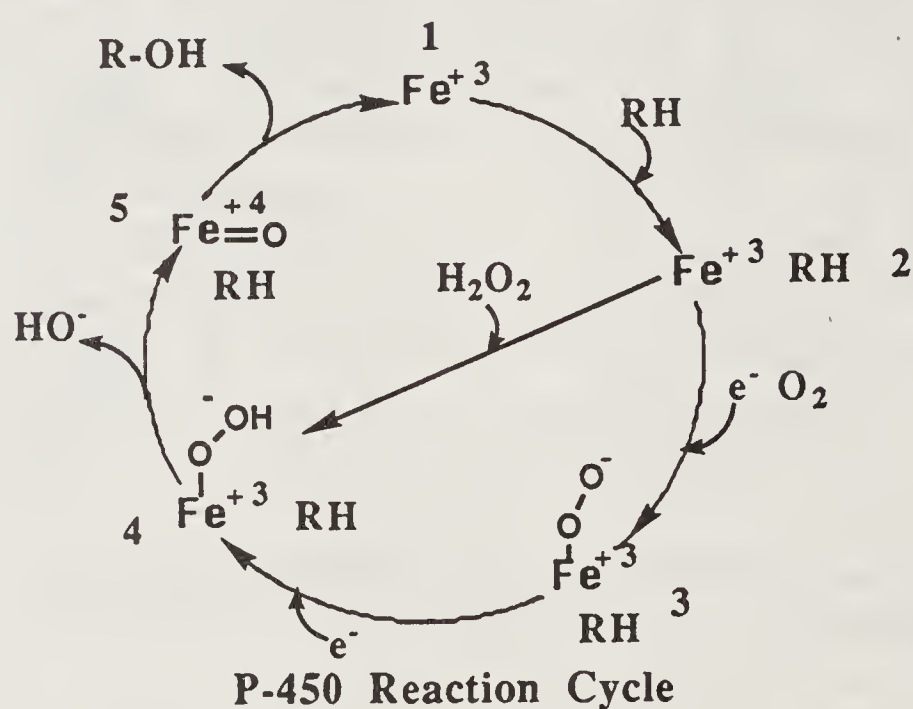


Figure 4. Stereoscopic view of the P-450cam active site. The carbonyl oxygen atom of the camphor molecule is darkened and the hydrogen bond between this group and the Tyr 96 OH is shown as a thin line.

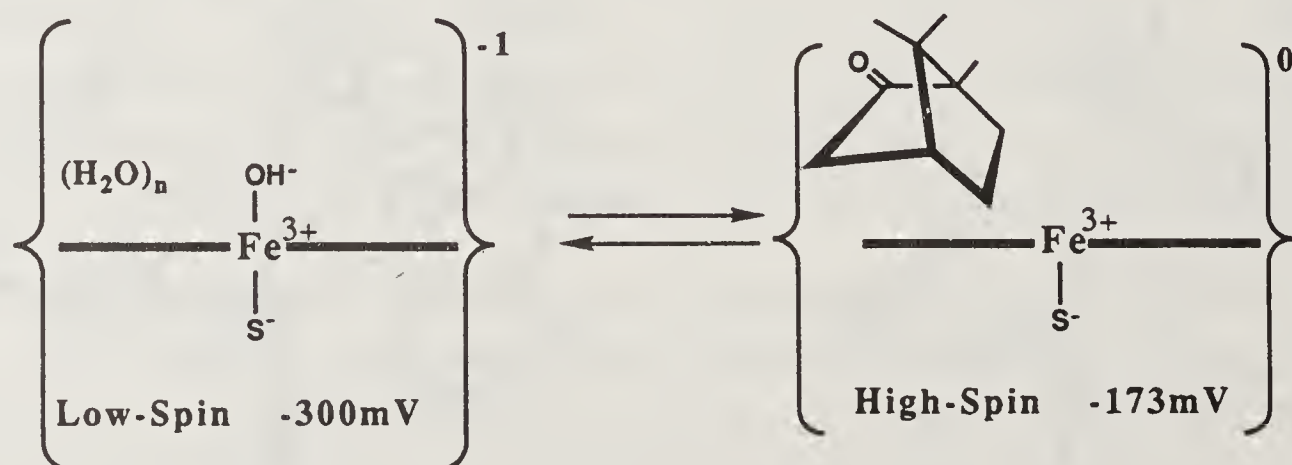
where it catalyzes the stereo- and regio-specific hydroxylation of camphor to give 5-*exo* hydroxy camphor (35).



The overall catalytic cycle for P-450 is outlined in the following scheme.



In the resting state, the enzyme is ferric, hexacoordinate, and low spin. When substrate (RH) enters the pocket it displaces several active site water molecules occupying the substrate pocket and the axial aqua ligand to give a pentacoordinate, high spin complex. This step also leads to an increase in redox potential from -300 mV to -173 mV (36, 37). The hydration property at the active site is the most likely explanation for these substrate-induced changes in redox and electronic states of the heme.



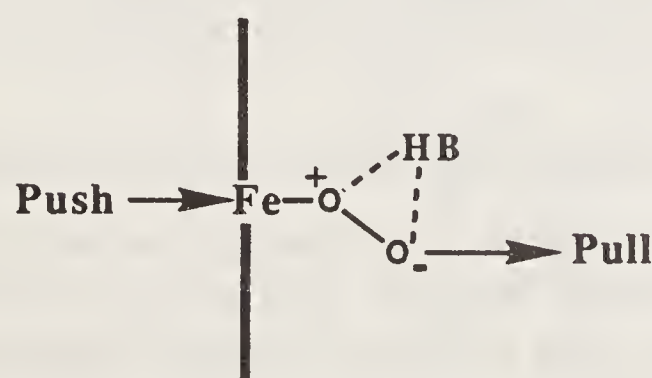
In the absence of substrate, the active site solvent tends to stabilize the ferric (Fe^{3+}) state and the excess -1 charge on the thiolate-iron-OH system when substrate binds thus leading to a low spin, low redox state. When camphor displaces these solvent

molecules, the net charge on the heme system becomes 0 and the local dielectric milieu decreases leading to an increase in redox potential and spin state.

Once the substrate is bound (2 of the P-450 reaction cycle) an iron-sulfur protein called putidaredoxin reduces P-450 cam from Fe^{3+} to Fe^{2+} thus enabling oxygen to bind to the ferrous heme. In microsomal P-450s, the electron donor is a flavoprotein. Intermediates 1, 2 and 3 in the reaction cycle have been well investigated using a variety of spectral methods. Crystal structures of 1 and 2 have been determined (3, 38) and the crystal structure of the carbon monoxy-substrate complex has served as a model for the oxy-complex (39). The remainder of the intermediates have not been identified and must exhibit transient lifetimes if they exist at all as distinct species. This is because the second electron transfer step giving hypothetical intermediate 4 in the P-450 reaction cycle is rate limiting and all one observes after this step is product. However, one reason for assuming an oxy-ferryl type intermediate after O–O bond cleavage is based on analogies with peroxidase. Moreover, it is possible for peroxide to substitute for oxygen and the two electron transfer steps (see P-450 reaction cycle). Model heme systems also support the idea that the O–O bond cleaves heterolytically as it does in peroxidase leaving behind a single oxygen atom which then proceeds to hydroxylate the substrate (40).

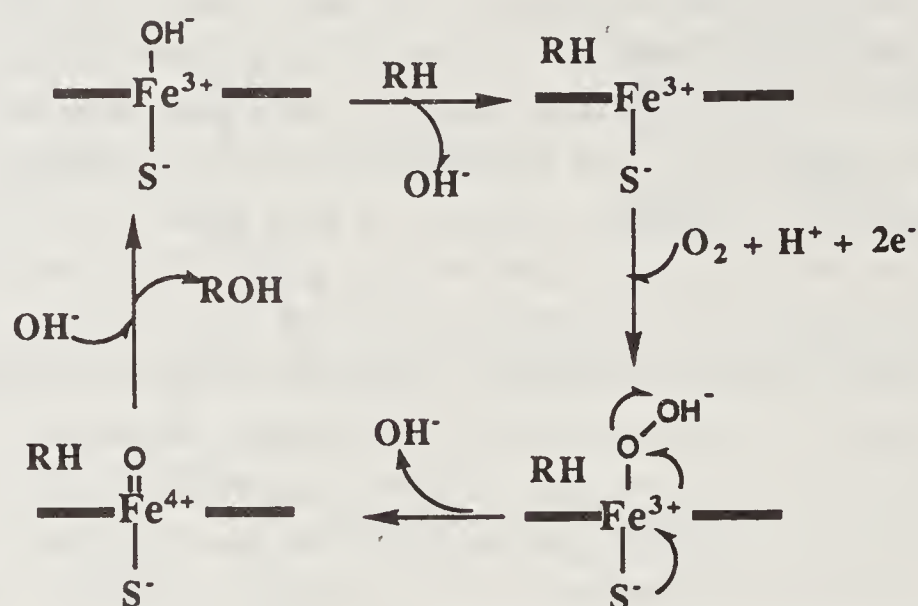
The next question is how P-450 cleaves the O–O bond and whether this process is similar to the way peroxidase works. Whatever the precise mechanism for P-450, the crystal structures provide clear evidence that P-450 cannot operate like a peroxidase. As shown in figure 4, the P-450 cam active site is formed by aromatic and aliphatic side chains and there is no group similar to the catalytic arginine and histidine in peroxidase. Indeed, there is no group that obviously can function as an acid–base catalyst or is suitably positioned to directly stabilize separated charges anticipated in the transition state of a heterolytic mechanism. Nor should this be expected. Model heme studies have shown that a protic environment similar to that found in peroxidase will promote iron oxidation and release of peroxide or water and will not allow for a stable ferrous–oxy complex (41,42). In this respect, P-450 more closely resembles the globins than peroxidase.

How, then, does P-450 cleave the O–O bond? The key very likely is the axial heme ligand. In P-450 the sulfur atom of a cysteine residue coordinates directly to the heme iron atom while in peroxidases and globins this residue is a histidine. The neutral-to-anion pK of cysteine is much lower than histidine and therefore cysteine is a better nucleophile than His. The sulfur of cysteine should be able to more effectively contribute electron density to the heme iron which in turn aids in cleaving the O–O bond. This has led to the “push–pull” concept for heme enzymes outlined below.

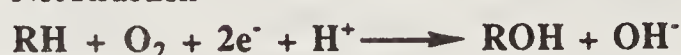


Since peroxidase has a relatively weak electron “push” on the heme iron, groups on the distal side are required to assist in cleaving the O–O bond: the acid–base catalyst

B (distal His) and suitably positioned charged groups that “pull” (i.e. the Arg in peroxidase). Because P-450 requires a hydrophobic distal pocket both for oxygen and aliphatic/aromatic substrate binding, a stronger “push” is required. This is the role of the cysteine ligand and was recognized as such before the X-ray structure became available (43). This leads to the P-450 mechanism outlined below.



Net Reaction



In this mechanism, the resting enzyme is low-spin and hexacoordinate with the Cys and a hydroxide or water molecule forming the axial coordination ligands. The substrate (RH) enters the pocket and displaces the axial aqua ligand resulting in a shift of the spin state from hexacoordinate low spin to pentacoordinate high spin. The stereochemistry of these two species has been determined from refined crystal structures (38, 4). The remaining two, the peroxy and ferryl intermediates, are hypothetical. Nevertheless, this scheme emphasizes the point that the sulfur of the cysteine ligand plays an essential role in cleaving the O–O bond.

One final difference between peroxidase and P-450 that should be stressed is in the mechanism of substrate oxidation. Although both P-450 and plant peroxidases oxidize organic substrates, P-450 does so by inserting the iron-linked oxygen atom into the substrate C–H bond while peroxidase does so by electron transfer. That is, the substrate is oxidized by Compound I to give a substrate radical which further reacts with other substrate molecules in a nonenzymatic process. The crystal structures show that one important reason for this difference is accessibility to the active site. Substrates are unable to approach closely to the oxy-ferryl intermediate in peroxidase so direct hydroxylation cannot occur. Instead, oxidation occurs at the heme edge which presupposes an aromatic binding site in this region. In contrast, P-450s possess a specific binding pocket for hydrocarbon substrates near the heme iron atom.

The P-450 O₂ Binding Site and Acid Catalysis

A serious omission in what has been discussed so far about the P-450 reaction cycle is the source of protons. In order for the O–O bond to cleave, the leaving oxygen atom must depart with a proton whether it leaves as a hydroxide ion or radical. From the X-ray structure the identity of the acid catalyst is not immediately obvious. However, recent work in our lab on the crystal structure of various substrate complexes (39) coupled with mutagenesis studies (44, 45), has provided some insights

into a possible route of protons into the active site. First, however, it is necessary to discuss what we know about the P-450 O_2 binding site.

Although P-450 cam shares only limited sequence homology with eukaryotic P-450s, it is clear that the central, distal helix which contacts both the substrate and oxygen is highly conserved and similarly positioned in all P-450s (4). Directly adjacent to the O_2 binding site, the distal helix experiences a local disruption in regular helical hydrogen bonding geometry (figure 5). This is because the carbonyl oxygen atom of Gly 248 accepts a hydrogen bond from the side chain OH group of the highly conserved Thr 252. Such a disruption forces a slight opening of the helix and produces a groove into which oxygen can bind. The refined crystal structure of the P-450 cam-substrate-carbon monooxy complex showed that the CO ligand is bent toward Thr 252 as expected from model building studies (46).

Not surprisingly, Thr 252 has been one of the early targets for site-directed mutagenesis studies. When Thr 252 is converted to Ala or Val, nearly all the electrons flowing into P-450 in a reconstituted assay system are funneled into the production of H_2O_2 and water while only 5% go to hydroxylation of substrate (44, 45). The Thr 252 to Ser mutant is much more efficient giving 0.81 moles of product per mole of oxygen consumed (44, 45) presumably because Ser can provide the same side chain hydrogen bond as can Thr. Thus, replacement of Thr 252 "uncouples" electron transfer from hydroxylation. Since the same uncoupling effect is observed when various substrates other than camphor are used, one is tempted to look for a common explanation for uncoupling. The crystal structure of several of these alternative

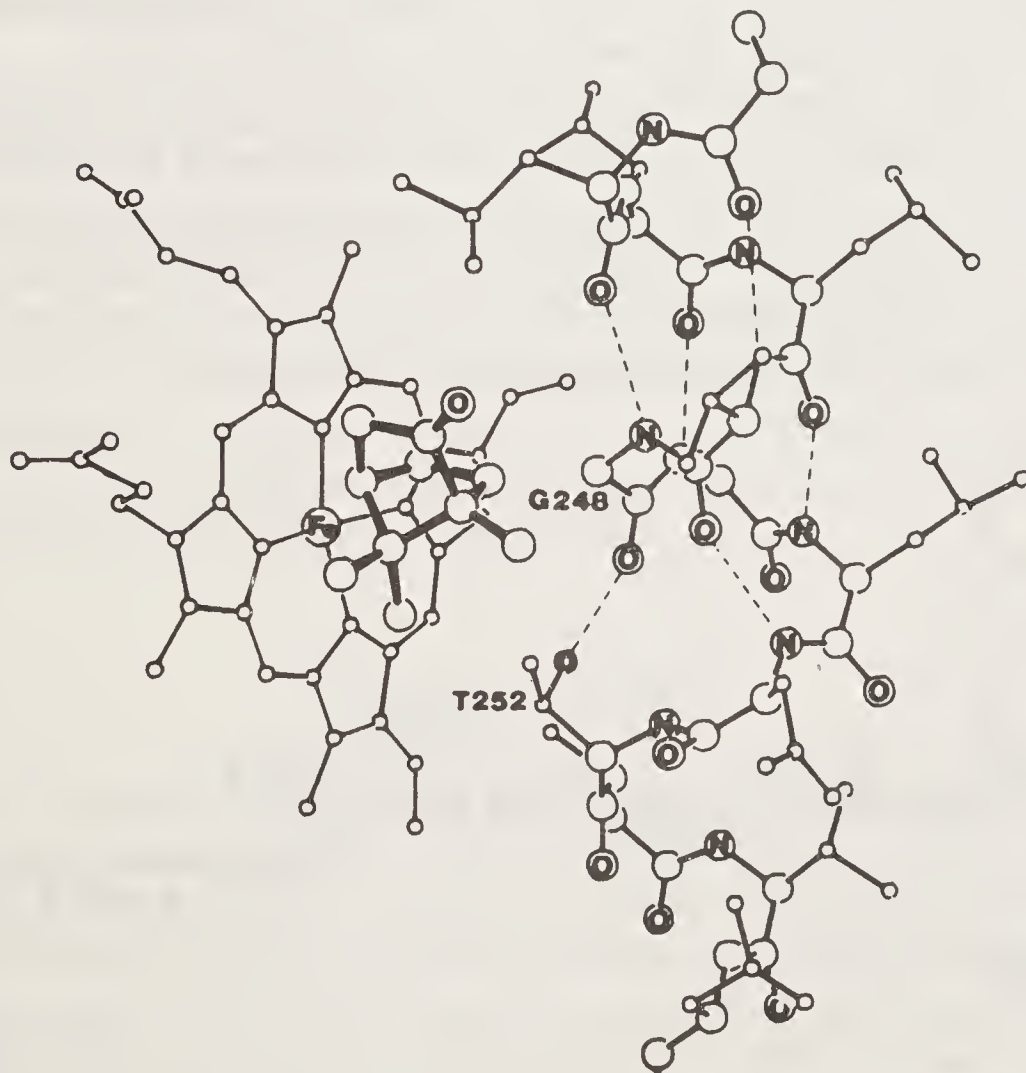
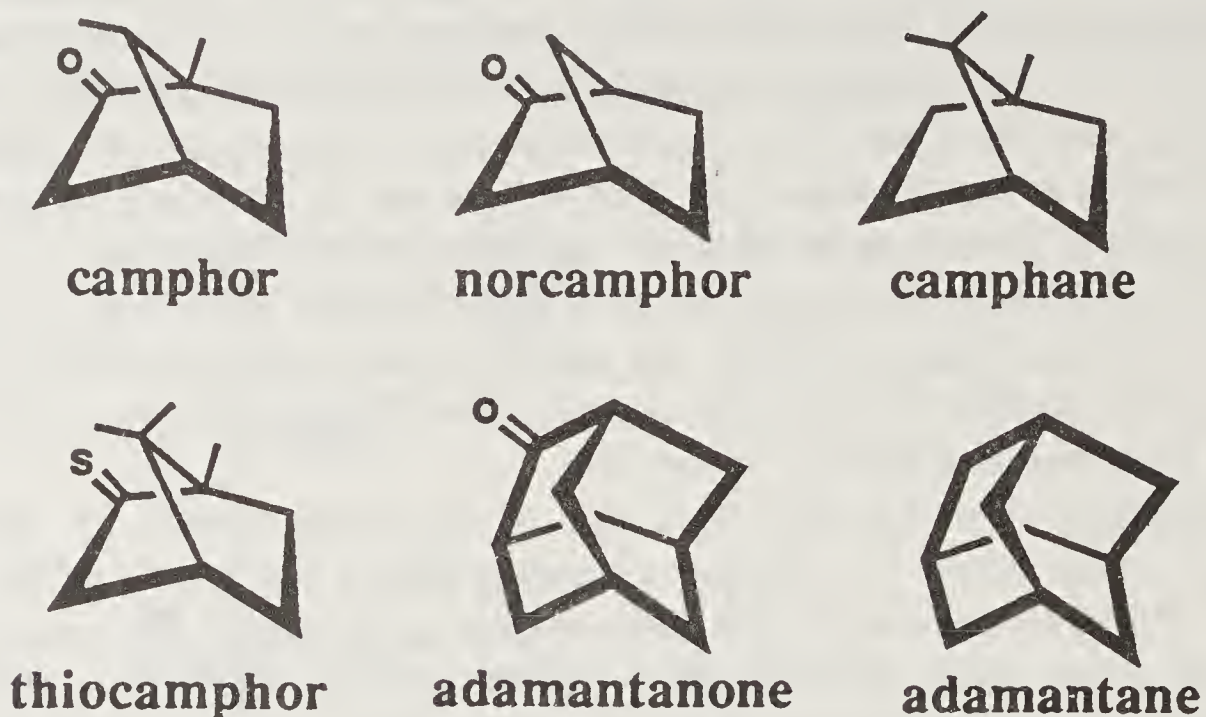


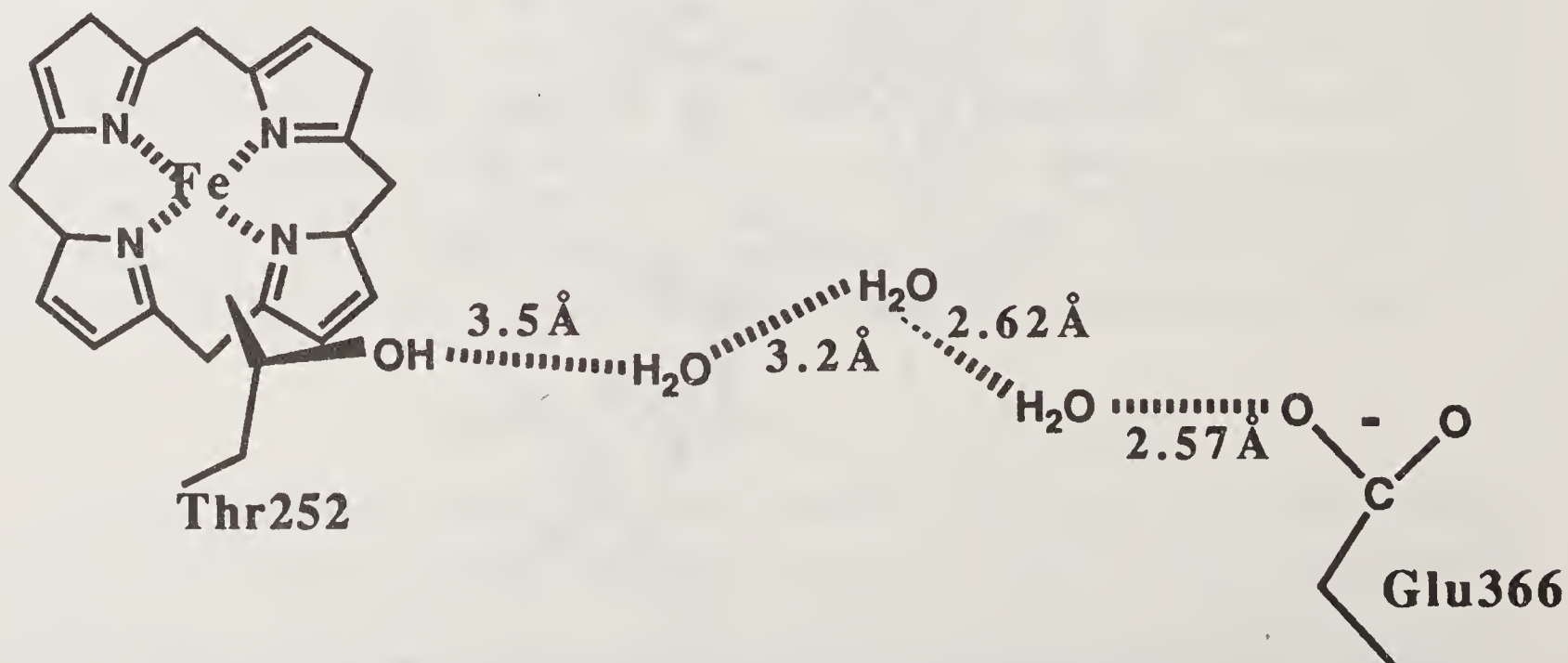
Figure 5. The highly conserved segment of the distal helix I in P-450 cam. Note that the carbonyl oxygen atom of Gly 248 accepts a hydrogen bond from the side chain OH of Thr 252 resulting in a local distortion of the local helical geometry.

substrate complexes have been refined to 2 Å resolution (39; Raag & Poulos, unpublished) and the structures of these substrates are provided below.



The general trend is that those substrates which are smaller than camphor or lack a carbonyl oxygen atom are uncouplers and not efficiently hydroxylated. The crystal structures show that those substrates which are uncouplers are hexacoordinate with a water molecule still coordinated to the iron atom while efficiently metabolized substrates like camphor and adamantanone are pentacoordinate. In addition, the uncoupling substrates are considerably more mobile in the active site as judged by crystallographic temperature factors. The greater degree of flexibility and the presence of the axial aqua ligand allows for greater penetration of solvent into the active site thus providing a protic environment that promotes the release of H_2O_2 .

How, then, are these observations related to Thr 252 and the oxygen binding site? We suggest that the local disruption of the distal helix due to Thr 252 also provides an access route for solvent to the iron-bound oxygen molecule. After examining many refined P-450cam structures, we found that Thr 252 is part of an internal, buried solvent channel consisting of ordered solvent molecules anchored between Thr 252 and Glu 366 as shown in the following schematic diagram.



Glu 366 is highly conserved and completely buried in the P-450cam structure. This buried channel seems to us a reasonable route for protons to the active site. If so,

the flow of solvent/protons must be controlled or else the reaction cycle will uncouple as it does when Thr 252 is converted to Ala or Val. These mutants may disrupt the channel making solvent more readily accessible to the iron-linked oxygen atom which in turn leads to peroxide release. This idea can be directly tested by determining the X-ray structure of Thr 252 mutants and perhaps by changing Glu 366 to other residues. Both experiments are being actively pursued.

Summary to Oxygen/Peroxide Activation

What has been discussed thus far illustrates how two proteins, peroxidase and P-450, have evolved two distinct and very different mechanisms for cleaving the O–O bond. A general view that emerges is that peroxidase seems to be a more sophisticated enzyme or at least satisfies our preconceived notions about how an enzyme ought to work. Peroxidase and catalase contain suitably positioned groups that assist in heterolytic cleavage of the peroxide O–O bond. The mechanism of storing oxidizing equivalents is complex. The enzyme active site is designed to produce relatively stable and long-lived oxy-ferryl and heme or amino acid centered radicals which then are reduced by substrates. P-450 shares none of these properties. There is no active site histidine or other neighboring charged group to assist in cleavage of the O–O bond. Nor is there a means of stabilizing a Compound I-like intermediate since although it might form, the lifetime is so short that this hypothetical intermediate has escaped detection. It appears that the main role of the protein is to provide a hydrophobic pocket for substrate binding and an appropriate milieu to allow thiolate–iron–oxygen activation chemistry to proceed. In this view, the protein plays a passive role as “solvent” rather than an active role as it does in peroxidase. There is one possible exception that may elevate P-450 to a more sophisticated status as an enzyme: control of solvent entry into the pocket. We have postulated the presence of a solvent channel in P-450 that controls the access of solvent and hence protons to the active site which is required for protonation of the oxygen ligand. Too much leads to peroxide release and decreased efficiency of hydroxylation. What is required is precisely the right number of protons at the right time. We know little about how this works although some such mechanism must be operating. This, then, presents a somewhat unorthodox catalytic role for the protein. Rather than the usual use of side chains in directly stabilizing the transition state, the protein controls solvent access to the active site during the catalytic cycle.

Electron Transfer

Activation of peroxides and oxygen is one part of a heme enzyme reaction. The second is electron transfer. For the plant peroxidases this involves oxidation of an organic substrate by Compound I. We know little about this process although we can make some educated guesses regarding the limits on stereochemistry. If the plant peroxidases resemble CCP which sequence alignments and homologous modeling suggest (47), the heme in plant peroxidases also is buried and not easily accessible. However, a solvent channel leads to the distal pocket although this pocket cannot allow close entry of an organic dye for direct interaction with the iron-linked oxygen atom of Compound I and hence no hydroxylation. Instead, the substrate is oxidized

at the heme periphery where it reduces the heme cationic radical. What remains to be determined is the nature of this substrate binding site at the heme periphery.

A more challenging problem and the one receiving increasing attention is electron transfer between proteins. CCP requires ferrous cytochrome c in order to reduce both Compounds I and II. As noted earlier, P-450cam is reduced by a ferredoxin-like (Fe_2S_2) protein called putidaredoxin. Irrespective of the specific proteins involved, there are two steps in this process. First, the proteins must form a complex and second the electron is transferred as indicated in the following reaction mechanism where D = donor and A = acceptor,



Protein-Protein Complexes

Most of the work on redox partner complexes have centered on cytochrome c although the principles derived from cytochrome c systems seem to be generally applicable to many others. Complexes formed between cytochrome c and its redox partners are highly dependent on ionic strength indicating that electrostatics plays an important role. It also is known that a ring of highly conserved lysines surrounding the exposed heme edge are involved in such complexes for both donors and acceptors. Thus, the electron path in and out in cytochrome c are one and the same.

There are only a few examples where the crystal structures of both redox partners are available. This includes CCP-cyt.c (11) and cyt.b₅-hemoglobin (48). There are, however, additional examples where the structures are known but the two proteins are not natural physiological partners although the partners do transfer electrons very efficiently. The best example is cyt.c-cyt.b₅. At one time these were the only two partners whose structures were known and this prompted the first attempt at generating a hypothetical model of a redox complex (49). In this model, the two exposed heme edges came very close (about 8 Å) which enables direct electron transfer between heme edges (figure 6).

Over the years, however, the physiologically relevant CCP-cyt.c complex has received considerable attention. The hypothetical model of this complex has served as a basis for experimentation and testing ideas about interprotein electron transfer. In principle this complex is not much different than the cyt.c-cyt.b₅ complex in that complementary charges dictate how the two molecules will bind to one another. The main difference is that in the CCP-cyt.c complex, the two heme edges are about 17 Å apart necessitating electron transfer *through* protein (figure 7).

On a crude level, the initial model proposed in 1980 (50) and refined in 1984 (11) has held up. Nevertheless, it has become increasingly clear that this static view of the model is misleading. First, the two surfaces of CCP and cyt.c while electrostatically complementary are not complementary in shape. This is an important difference. Electrostatic energy will fall off as $1/r$ where r is the distance between electrostatic centers. If we view CCP and cyt.c as two oppositely charged electrostatic plates, then a slight reorientation of these two plates will not cost much in electrostatic energy and hence the stability of the complex. On the other hand, nonbonded and hydrophobic forces controlled by complementary shape interactions as in antibody-protein antigen complexes fall off as $1/r^6$. Thus, a slight repositioning costs a great deal in terms of free energy of binding. One might also expect severe restrictions on

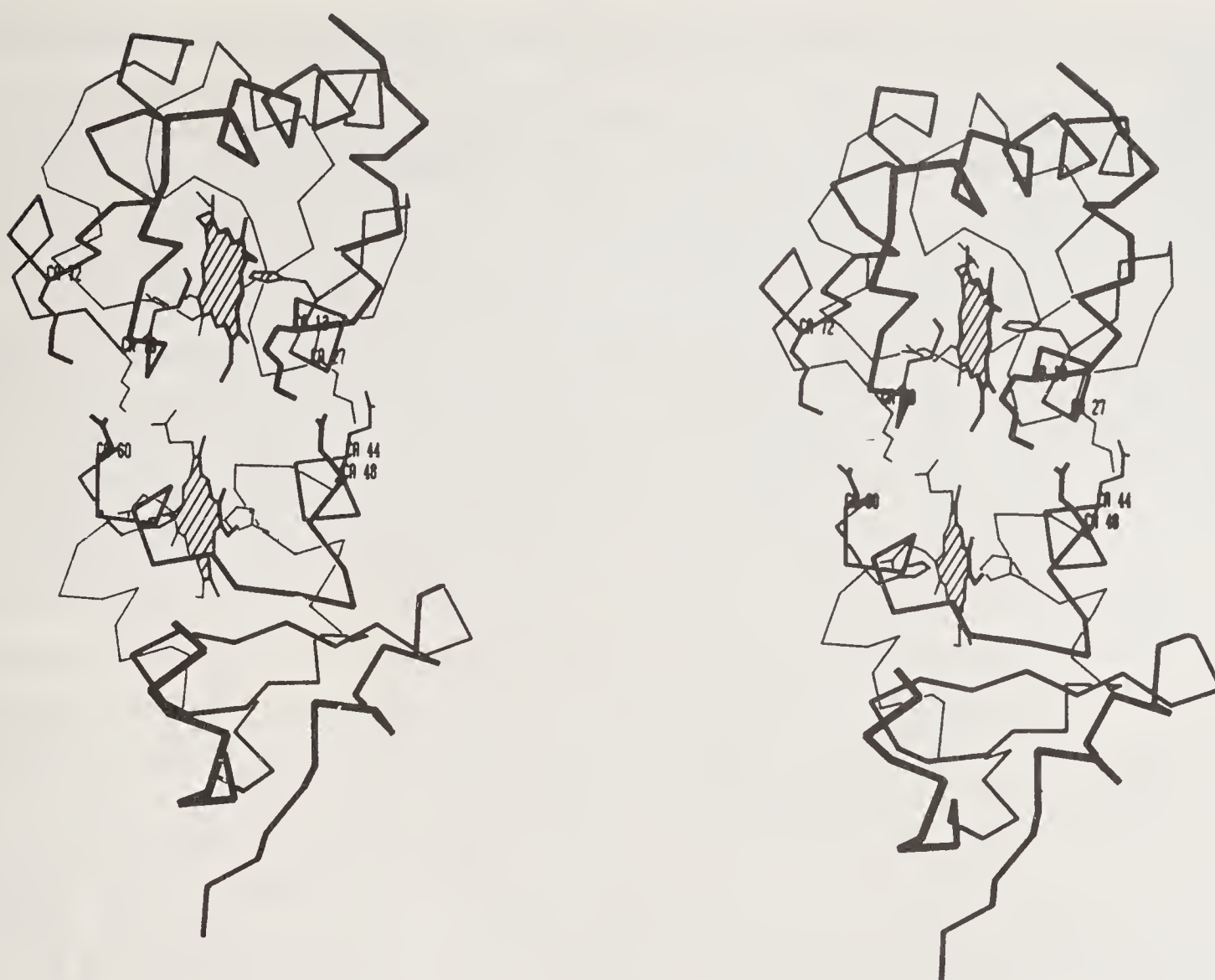


Figure 6. Stereoscopic view of the cyt.c-cyt.b5 electron transfer complex based on the model first proposed by Salemme (1976).

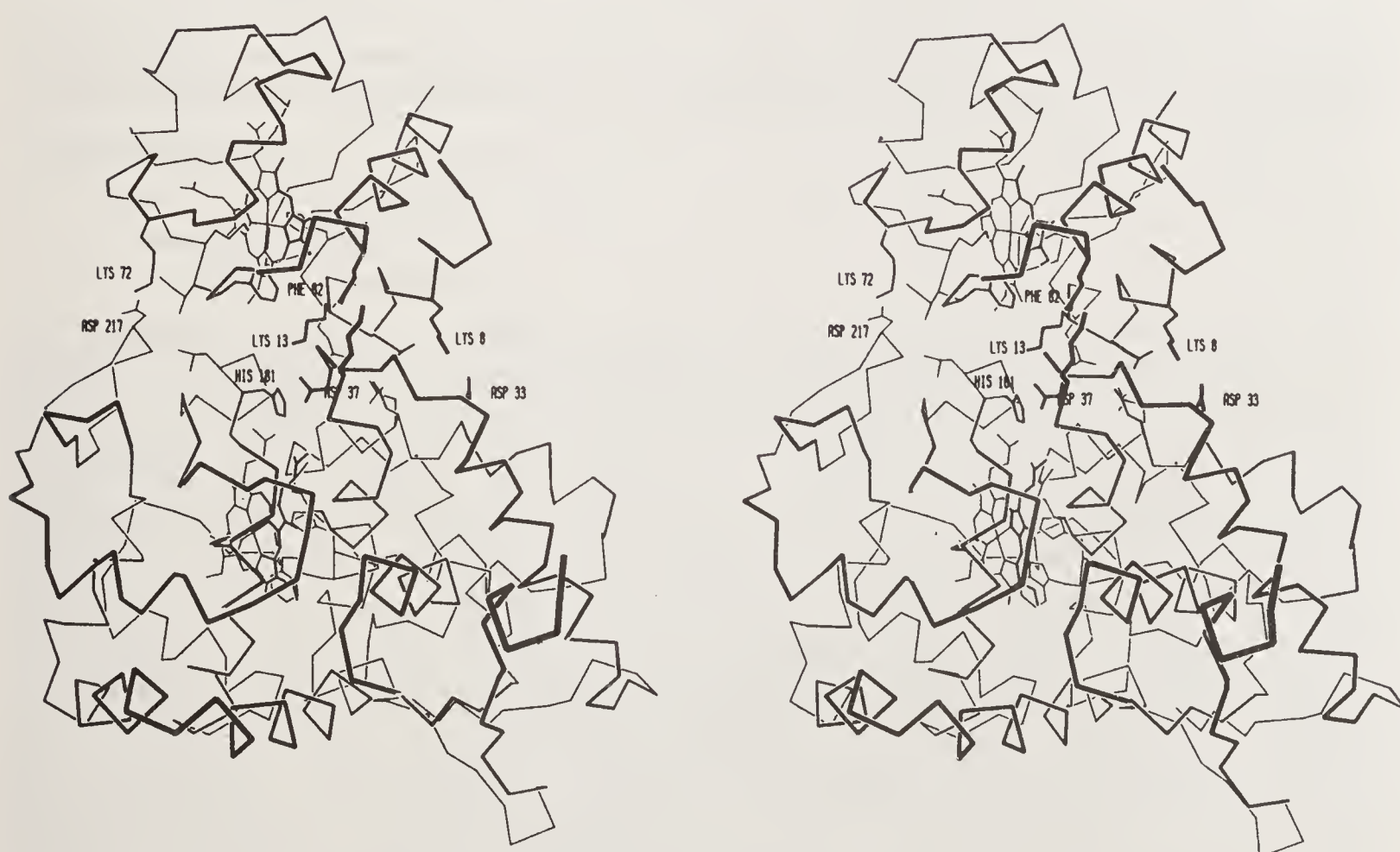


Figure 7. Stereoscopic view of the hypothetical CCP-cyt.c complex. Key ionic groups forming potential electrostatic contacts are shown. Also note the position of Phenylalanine 82 in cyt.c and His 181 in CCP.

orientation if specific hydrogen bonds contribute significantly to the free energy of binding.

There are both theoretical and experimental reasons to believe that the CCP–cyt.c recognition does not result in a single, unique complex. Northrup *et al.* (51) have carried out a Brownian dynamics study of the complex and find additional sites on the surface of CCP to which cyt.c should be able to bind. The most highly populated site is the one predicted on the basis of our hypothetical modeling but there are two other regions on CCP that the calculations indicate could serve as successful docking sites. The experimental evidence against a static model is also mounting. McLendon *et al.* (52) found that the association constant between CCP and cyt.c determined by fluorescence and NMR methods gave very different values ranging from 10^8 M^{-1} to 10^4 M^{-1} . This indicates that there must be more than one form of the complex.

A totally different experiment leads to similar conclusions. Tollin and coworkers have worked out a laser flash photolysis method that allows one to directly measure the rate of electron transfer from ferro cyt.c to CCP Compound I (53). In this method, ferric cyt.c, CCP Compound I, and a flavin are mixed together in solution. The entire system is stable in the absence of light. Upon exposure to a laser flash, the flavin is photoreduced which rapidly delivers an electron to cyt.c giving ferro cyt.c. What is monitored is the rate of cyt.c reduction by the flavin and/or the first order rate of electron transfer from ferro cyt.c to Compound I in the 1:1 complex. One of the more interesting findings from these studies is that different species of cyt.c give different rates of reduction of the cyt.c heme by the flavin when complexed with CCP but not when free in solution. These results indicate that the degree of heme accessibility is not the same in the different CCP–cyt.c complexes further implying that CCP does not form exactly the same complex with different cytochromes c (53).

Finally, we have obtained indirect crystallographic evidence that the complex is heterogeneous. It has been possible to cocrystallize CCP with horse heart cyt.c at low ionic strength and the structure was solved using molecular replacement techniques (54). Unexpectedly we found electron density for CCP but none for cyt.c. The crystalline lattice consisted of CCP dimers separated by very large “empty” regions big enough to accommodate cyt.c. SDS gel analyses clearly showed the presence of cyt.c in the crystals so we must conclude that the cyt.c is present but so severely disordered that the cyt.c molecules do not significantly contribute to the diffraction pattern. The disorder could be either dynamic or static. If dynamic, the cyt.c could be freely rotating in the crystalline lattice or “wobbling” by several angstroms. If static, the cyt.c could be fixed in several different orientations. In either case, these data lend some support to the notion that a single unique complex is probably too naive a view. In hindsight, it appears that using low ionic strength for crystallization may have been the wrong approach. Although the association constant goes up at low ionic strength, the rate of electron transfer goes up with *increasing* ionic strength (55). What may be happening is that at the higher ionic strength, the main productive complex is favored with less competition from slightly weaker, less productive complexes. Thus, by crystallizing from low ionic strength solutions, we may have encouraged the formation of multiple, nonproductive complexes. The general view that emerges from these various approaches favors a dynamical model with the hypothetical CCP–cyt.c complex representing a snapshot of one conformer and very likely a dominate conformer. Therefore, the model can be used as a guide to addressing questions regarding interprotein electron transfer albeit very cautiously.

To briefly summarize, in intermolecular electron transfer complexes electrostatics plays an important role in controlling recognition. Little is known about whether such complexes result in important conformational changes although the spectroscopic evidence to date argues against large changes. Based on the CCP–cyt.c system, the 1:1 complexes while exhibiting dissociation constants in the micromolar range, may not form a unique intermolecular complex but either several distinct complexes or a dynamical range of motions about some primary recognition site. This raises several questions. For example, are all complexes competent in electron transfer or must one particular orientation of the donor and acceptor be sampled before the electron transfer event is triggered? Because of the requirement or at least our belief that such systems must maintain a high degree of specificity, we favor the view that a limited few and probably only one recognition site is competent in electron transfer.

Electron Transfer Mechanisms

Once the complex is formed, the electron must transfer through protein unless one invokes a rather massive and very unlikely conformational change which allows the two hemes to directly contact one another. This problem of electron transfer paths in proteins is both fascinating and elusive.

There have been a variety of imaginative approaches to this problem but all begin with the basic tenets of Marcus theory (56). The theory attempts to explain the relationship between thermodynamic driving force and various reorganization energy terms to the rate of electron transfer. If we assume simple transition state theory for electron transfer, the rate of transfer is given by

$$k_{et} = \kappa \nu \exp(-\Delta G^*/RT)$$

where ΔG^* is the activation free energy, κ is the probability of electron transfer and ν is the frequency of motion along the reaction coordinate. The activation free energy ΔG^* is given by

$$\Delta G^* = \frac{(\Delta G^0 + \lambda)^2}{4\lambda}$$

where λ is the reorganization energy and ΔG^0 is the free energy change of the reaction and usually is taken as the difference in redox potential between donor and acceptor. λ is the energy required to alter bonds about the metal centers and reorganize the solvent. The theory predicts that the reaction rate will *not* continue to increase with increasing driving force but will plateau when $\Delta G^0 = -\lambda$. This provides a powerful test of the theory if all other factors except the driving force can be held constant. As might be expected, this becomes very difficult to apply to proteins since one does not have easy control over the various parameters contributing to the rate of electron transfer. However, some advances have been made because it is possible to replace the redox centers of certain heme proteins thereby altering the redox potential without altering the recognition surface. Site-directed mutagenesis offers another approach in modulating electron transfer parameters.

One means of altering the driving force of a reaction is to change the metal ion. Fortunately, it is possible with many heme proteins to substitute the heme with porphyrins containing other metal ions. The metal-replacement method as a means

of altering the driving force has been used successfully for the cyt.c/cyt.b5 system and it was found to follow Marcus theory (57). That is, a plot of rate of electron transfer vs. driving force showed an inverted region. Similar experiments with the CCP/cyt.c system also obey Marcus theory if one assumes a large reorganizational energy (1.5 eV) although not enough data were obtained to clearly discern the inverted region (58). These experiments provide information about the reorganization energy and give a qualitative estimate of the degree of changes that occur around the metal centers and in reorganizing the solvent.

Overall, the results are encouraging and give one confidence that proteins conform to conventional Marcus theory. If, however, the rate of electron transfer does not change with driving force, some other process is rate limiting. Hoffman and coworkers have termed such a process conformational gating (59). That is, some structural property of the complex is rate limiting. McLendon *et al.* recently have found evidence for conformational gating in the cyt.c/cyt.b2 system (60). The same metal replacement procedure showed no change in rates with driving force clearly implicating some other process not directly related to Marcus theory as rate limiting. For the experimentalist this is an attractive development since one ought to be able to change protein structural components using modern mutagenesis methods.

Finally is the question of path of electron transfer. Does the protein provide a specific set of connections between donor and acceptor redox centers? Providing such a circuit for electron transfer is an attractive idea since this is one means of controlling specificity. The question is how to attack this problem experimentally. Two methods are currently being used. The first is based on pioneering work on the use of ruthenated heme proteins to measure intramolecular rates. Surface histidines on proteins can specifically be labeled with $\text{Ru}(\text{NH}_4)_5^{3+}$ compounds (61, 62). Since the ruthenium exhibits redox potentials near those of many heme proteins, it is possible to measure the first order rate of electron transfer between the surface ruthenium and heme iron. If the chemistry of ruthenation can be controlled such that only one histidine is modified, then one can correlate not only the driving force but the distance between metal centers with rates. This has proved very useful in measuring ruthenium/iron intramolecular rates in ruthenated cytochrome c and myoglobin (63). At present this is the only method available for specifically and predictably changing the distance between redox centers although one is limited by where the histidine residues occur. With modern mutagenesis methods, however, one can place the histidine to be modified anywhere on the surface to vary the distance between redox centers. A second obvious approach is to use mutagenesis to change the protein between the redox centers.

The data available on natural redox partners are less extensive but again focus on the CCP/cyt.c system. In the hypothetical CCP/cyt.c system, the invariant Phe 82 of cyt.c and His 181 of CCP are close (figure 7). His 181 forms a hydrogen bonded network that connects the surface of CCP with the distal heme pocket (figure 2) while Phe 82 contacts the cyt.c heme (figure 7). Both have been mentioned as possible mediators of the electron transfer reaction. Both residues have been converted to other side chains. Mutagenesis has shown that in the steady state reaction, Phe 82 is not essential (64, 65). Nevertheless, there is always the question in steady state systems as to the nature of the rate limiting step. Electron transfer may be quite fast and not limiting in which case steady state kinetics measures some other limiting process. However, Hoffman and coworkers (66) have developed a system which

potentially overcomes this problem. Again metal substitution is used where the CCP heme iron is replaced with zinc. First, the zinc in the CCP:cyt.c complex is photo-excited to the triplet state where it reduces cyt.c(Fe^{3+}) to cyt.c(Fe^{2+}) giving a CCP zinc protoporphyrin π cation radical. In the second reaction, cyt.c(Fe^{2+}) reduces the CCP zinc protoporphyrin π cation radical. This last reaction is the one of interest and it was found that replacing Phe 82 with a serine or glycine leads to a large drop in the rate of reduction of the CCP zinc π cation radical but not when tyrosine is used in place of Phe 82. These results suggest that an aromatic residue at position 82 is important for efficient electron transfer. Thus, the steady state kinetics and the excited zinc system give very different results. As noted earlier, one could argue that in the steady state system, electron transfer is not rate limiting and that some other process is rate limiting so the zinc CCP:cyt.c system is a better estimate of the true first order rate of electron transfer. However, caution must be exercised here since the zinc system is far from the true physiological reaction. In particular, as noted by Pielak *et al.* (65), the high energy of the zinc protoporphyrin π cation radical could lead to extensive delocalization of spin density over aromatic centers that are not involved in the physiologically relevant reaction.

His 181 of CCP also has been a target for mutagenesis. His 181 hydrogen bonds with one of the heme propionates (figure 2) and provides a hydrogen bonded conduit from the surface to the distal pocket. Years ago Asakura and Yonetani (67) found that methylation of the heme propionates in CCP does not alter the reaction with peroxide to give Compound I but does greatly inhibit the electron transfer reaction. Since methylation would destroy the His 181-propionate hydrogen bond, we proposed that this hydrogen bonded network is important for electron transfer. This argument was further strengthened by the hypothetical model of the complex where His 181 was found to be near the center of the intermolecular interface (11) and also by chemical modification data which supported the model (68, 69). Yet replacing His 181 with Gly has very little effect on the crystal structure and reduces the steady reaction with cyt.c and the first order transfer of an electron from cyt.c(Fe^{2+}) to CCP Compound I by only a factor of two (70). These modest effects argue against some special role for His 181. Nevertheless, a Gly at position 181 would not disrupt the ability of the heme propionate to hydrogen bond with solvent. Thus, solvent could replace His 181 in the intermolecular complex and provide a hydrogen bonded network. The obvious experiment here is to replace His 181 with residues like Leu or Ile which cannot hydrogen bond and should significantly change the hydrogen bonding capabilities of the propionate. The role of His 181 remains unsettled.

Perhaps the strongest evidence to date for the role of a specific residue in electron transfer is the Trp 191 to phenylalanine mutant in CCP (30). This mutation does not effect the rate of reaction with peroxide to give Compound I but nearly eliminates the electron transfer reaction. As discussed earlier, the Trp 191 region is peculiar since the position and crystallographic temperature factors for Trp 191 are sensitive to the state of ligation at the heme iron atom (33). This argues for "communication" between the state of the heme iron and Trp 191. In addition, Trp 191 is situated at the C-terminal end of a surface loop connecting the heme ligand, His 175, and Trp 191 on the interior of the protein. Note that His 181 situated at the CCP–cyt.c intermolecular interface is between these two residues.

At present, however, these results are interesting observations that have primarily served to focus our attention on particular regions of the protein. Most of these results

are very recent and the field of biological electron transfer is just beginning to bring the tools of modern molecular biology coupled with some sophisticated chemistry to bear on the problem. The future looks exciting.

Acknowledgements

I would like to thank Dr. Matthew Mauro and Dr. Steven Edwards for useful comments and suggestions.

References

1. Poulos, T. L., Freer, S. T., Alden, R. A., Edwards, S. L., Skoglund, U., Takio, K., Eriksson, B., Xuong, N., Yonetani, T. & Kraut, J. (1980) *J. Biol. Chem.* **255**, 575–580.
2. Murthy, M. R. N., Reid, T. J., Sicignano, A., Tanaka, N. & Rossmann, M. G. (1981) *J. Mol. Biol.* **152**, 465–499.
3. Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C. & Kraut, J. (1985) *J. Biol. Chem.* **260**, 16122–16130.
4. Poulos, T. L., Finzel, B. C. & Howard, A. J. (1987) *J. Mol. Biol.* **195**, 687–700.
5. Barlow, C. H., Maxwell, J. C., Wallace, W. J. & Caughey, W. S. (1973) *Biochem. Biophys. Res. Commun.* **55**, 91–95.
6. Koster, A. S. (1975) *J. Chem. Phys.* **63**, 3284–3286.
7. Demma, L. S. & Salhany, J. M. (1977) *J. Biol. Chem.* **252**, 1226–1231.
8. Schonbaum, G. R. & Lo, S. (1972) *J. Biol. Chem.* **247**, 3353–3360.
9. Finzel, B. C., Poulos, T. L. & Kraut, J. (1984) *J. Biol. Chem.* **259**, 13027–13036.
10. Fita, I. & Rossmann, M. G. (1985) *J. Mol. Biol.* **185**, 21–37.
11. Poulos, T. L. & Finzel, B. C. (1984) in *Peptide and Protein Reviews*, ed. Hearn, M.T.W. (Marcel Dekker, Inc., New York & Basel), Vol. 4, pp. 115–171.
12. Poulos, T. L. (1987) *Advances in Inorganic Biochemistry* **7**, 1–36.
13. Lang, G., Spartalian, K. & Yonetani, T. (1976) *Biochim. Biophys. Acta* **451**, 250–258.
14. Chance, M., Powers, L., Poulos, T. L. & Chance, B. (1986) *Biochemistry* **25**, 1266–1270.
15. Edwards, S. L., Xuong, N-h., Hamlin, R. C. & Kraut, J. (1987) *Biochemistry* **26**, 1503–1511.
16. Dolphin, D., Forman, A., Borg, D. C., Fajer, J. & Felton, R. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 614–618.
17. Yonetani, T. & Ray, G. S. (1965) *J. Biol. Chem.* **240**, 4503–4508.
18. Yonetani, T. & Ray, G. S. (1966) *J. Biol. Chem.* **241**, 700–706.
19. Coulson, A. F. & Yonetani, T. (1972) *Biochem. Biophys. Res. Commun.* **49**, 391–398.
20. Yonetani, T. (1976) in *The Enzymes*, ed. Boyer, P. D. (Academic Press, New York), Vol. 13, pp. 345–361.
21. Poulos, T. L. & Kraut, J. (1980) *J. Biol. Chem.* **255**, 8199–8205.
22. Erman, J. E., Vitello, L. B., Mauro, J. M. & Kraut, J. (1989) *Biochemistry* **28**, 7992–7995.
23. Hoffman, B. M., Roberts, J. E., Kang, C. H. & Margoliash, E. (1981) *J. Biol. Chem.* **256**, 6556–6564.
24. Myers, D. & Palmer, G. (1985) *J. Biol. Chem.* **260**, 3887–3890.
25. Hori, H. & Yonetani, T. (1985) *J. Biol. Chem.* **260**, 349–355.
26. Glotz, S., Kaput, J. & Blobel, G. (1982) *J. Biol. Chem.* **257**, 11186–11190.
27. Kaput, J., Glotz, S. & Blobel, G. (1982) *J. Biol. Chem.* **257**, 15054–15058.
28. Goodin, D. B., Mauk, A. G. & Smith, M. (1987) *J. Biol. Chem.* **262**, 7119–7124.
29. Fishel, L. A., Villafranca, J. E., Mauro, J. M. & Kraut, J. (1987) *Biochemistry* **26**, 351–360.
30. Mauro, J. M., Fishel, L. A., Hazzard, J. T., Meyer, T. E., Tollin, G., Cusanovich, M. A. & Kraut, J. (1988) *Biochemistry* **27**, 6243–6256.
31. Sivaraja, M., Goodin, D. B., Mauk, A. G., Smith, M. & Hoffman, B. M. (1989) *Science* **245**, 738–740.

32. Edwards, S. L., Kraut, J. & Poulos, T. L. (1988) *Biochemistry* **27**, 8074–8081.
33. Edwards, S. L. & Poulos, T. L. (1990) *J. Biol. Chem.* **265**, 2588–2595.
34. Yonetani, T., Yamamoto, H., Erman, J. E., Leigh, J. S. & Reed, G. H. (1972) *J. Biol. Chem.* **247**, 2447–2455.
35. Gunsalus, I. C., Meeks, J. R., Lipscomb, J. D., Debrunner, P. G. & Mucnk, E. (1974) in *Molec. Mech. of Oxygen Activation*, ed. Hayishi (Academic Press, New York), pp. 559–613.
36. Sligar, S. G. & Gunsalus, I. C. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1078–1082.
37. Sligar, S. G. & Murray, R. I. (1986) in *Cytochrome P-450: Structure, Mechanism, and Biochemistry*, ed. Ortiz de Montellano, P. (Plenum, New York), pp. 429–503.
38. Poulos, T. L., Finzel, B. C. & Howard, A. J. (1986) *Biochemistry* **25**, 5314–5322.
39. Raag, R. & Poulos, T. L. (1989) *Biochemistry* **28**, 917–922.
40. McMurray, T. J. & Groves, J. T. (1986) in *Cytochrome P-450: Structure, Mechanism, and Biochemistry*, ed. Ortiz de Montellano, P. (Plenum, New York), pp. 1–28.
41. Brinigar, W. S., Chang, C. K., Geibel, J. & Traylor, T. G. (1974) *J. Am. Chem. Soc.* **96**, 5597–5599.
42. Malstrom, B. O. (1982) *Annu. Rev. Biochem.* **51**, 21–59.
43. Dawson, J. H., Holm, R. H., Trudell, J. R., Barth, G., Linder, R. E., Bunnenberg, E. & Djerassi, C. (1976) *J. Am. Chem. Soc.* **98**, 3707–3709.
44. Imai, M., Shimada, H., Watanabe, Y., Matsushima-Hibiya, Y., Makino, R., Koga, R., Horiuchi, T. & Ishimura, Y. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7823–7827.
45. Martinis, S. A., Atkins, W. M., Stayton, P. S. & Sligar, S. G. (1989) *J. Am. Chem. Soc.* **111**, 9252–9253.
46. Raag, R. & Poulos, T. L. (1989b) *Biochemistry* **28**, 7586–7592.
47. Welinder, K. G. & Norskov-Lauritsen, L. (1986) in *Molec. and Physiolog. Aspects of Plant Peroxidases*, eds. Greeppin, H., Penel, C., & Gaspar, Th., University de Geneve-Centre de Botanique, pp 61–70.
48. Poulos, T. L. & Mauk, G. (1983) *J. Biol. Chem.* **258**, 7369–7373.
49. Salemme, F. R. (1976) *J. Mol. Biol.* **102**, 563–568.
50. Poulos, T. L. & Kraut, J. (1980) *J. Biol. Chem.* **255**, 10322–10330.
51. Northrup, S. H., Boles, J. O. & Reynolds, J. C. (1988) *Science* **241**, 67–70.
52. McLendon, G. L., Magner, E., Zhang, Q. P., Pardue, K. & Simmons, J. (1988) in *Robert A. Welch Found. Conf. on Chemical Res. XXXII Valency Chapter VII*, pp. 201–217.
53. Hazzard, J. T., Poulos, T. L. & Tollin, G. (1987) *Biochemistry* **26**, 2836–2848.
54. Poulos, T. L., Sheriff, S. & Howard, A. J. (1987) *J. Biol. Chem.* **262**, 13881–13884.
55. Hazzard, J. T., McLendon, G., Cusanovich, M. & Tollin, G. (1988) *Biochem. Biophys. Res. Commun.* **151**, 429–443.
56. Marcus, R. A. & Sutin, N. (1985) *Biochim. Biophys. Acta* **811**, 265–322.
57. McLendon, G. & Miller, J. R. (1985) *J. Am. Chem. Soc.* **107**, 7811–7816.
58. Conklin, K. T. & McLendon, G. (1988) *J. Am. Chem. Soc.* **110**, 3345–3350.
59. Hoffman, B. M. & Ratner, M. A. (1987) *J. Am. Chem. Soc.* **109**, 6234–6237.
60. McLendon, G., Pardue, K. & Bak, P. (1987) *J. Am. Chem. Soc.* **109**, 7540–7541.
61. Yocum, K. M., Winkler, J. R., Nocera, D. G., Bordignon, E. & Gray, H. B. (1983) *Chemica Scripta* **21**, 29–33.
62. Isied, S. S., Kuehn, C. & Worosila, G. (1984) *J. Am. Chem. Soc.* **106**, 1722–1726.
63. Mayo, S. L., Ellis, W. R., Crutchley, R. J. & Gray, H. B. (1986) *Science* **233**, 948–952.
64. Pielak, G. J., Mauk, A. G. & Smith, M. (1985) *Nature (London)* **313**, 152–154.
65. Pielak, G. J., Concar, D. W., Moore, G. R. & Williams, R. J. P. (1987) *Protein Engineering* **1**, 83–88.
66. Liang, N., Pielak, G. J., Mauk, G. A., Smith, M. & Hoffman, B. M. (1986) *Proc. Natl Acad. Sci. U.S.A.* **240**, 311–313.
67. Asakura, T. & Yonetani, T. (1969) *J. Biol. Chem.* **244**, 4573–4579.
68. Waldmeyer, B., Bechtold, R., Bosshard, H. R. & Poulos, T. L. (1982) *J. Biol. Chem.* **257**, 6073–6076.
69. Bosshard, H. R., Banziger, J., Hasler, T. & Poulos, T. L. (1984) *J. Biol. Chem.* **259**, 5683–5690.
70. Miller, M. A., Hazzard, J. T., Mauro, J. M., Edwards, S. L., Simons, P. C., Tollin, G. & Kraut, J. (1988) *Biochemistry* **27**, 9081–9088.

Conformational Changes at the B Chain N Terminus in the Insulin Hexamer

G. G. Dodson, R. E. Hubbard and B. Xiao

Department of Chemistry, University of York, York, YO1 5DD, UK

Our understanding of how protein molecules are constructed has been immeasurably helped by the recognition by Ramachandran and his colleagues at Madras that the dihedral angles assumed by peptides are limited to certain values (1). These permitted angles have been identified and are simply characterised in the widely used Ramachandran diagram. When peptides in conformations defined by the most energetically favourable dihedral angles are repeated periodically, they can form stable structures (the helix and sheet) which satisfy their main chain hydrogen bonding potential. These so-called secondary structures are the basic elements in protein construction. Their assembly together gives proteins their tertiary structure. Thus the relationship between dihedral angles and peptide and protein conformation is an important component in the study of proteins and their structural behaviour.

Insulin, like many proteins, exhibits both helix and sheet structures. The molecule consists of two peptide chains, the A chain of 21 amino acids and the B chain of 30 residues, with disulphide bonds between A6–A11, A7–B7 and A19–B20. It undergoes self-assembly to dimers which in the presence of zinc ions further assemble to hexamers, as illustrated in figure 1. The determination of the hexamer structure in the 2Zn insulin crystal in 1969 showed that it is organised as three dimers, each of which contains a local axis of 2-fold symmetry perpendicular to the 3-fold axis (2). These 2-fold axes intersect the 3-fold axis yielding a hexamer with 32 symmetry. The 2 zinc ions lie on the central 3-fold axis and are disposed 8Å away from the point of intersection of the three 2-fold axes. Each zinc ion is octahedrally coordinated to the imidazole N^ε of three symmetry related B10 His side chains, and to three water molecules. This arrangement means that each of the three dimers is bonded to the two axial zinc ions which hold the hexamer together.

X-ray analysis of the dimeric and monomeric insulins further reveals that the assembly to dimer or hexamer does not affect in any significant way the monomer's internal structure or much of its external surfaces (3). There were however different conformations observed at the N terminal B chain (residues B1–B8) which are not involved in dimer formation and which in that structure make few intramolecular contacts. Thus it appears that the insulin molecule's evolution has endowed the monomer with stable surfaces through which efficient self-assembly is achieved.

Crystallographic studies on the insulin hexamer over the last twenty years have revealed that it can accommodate substantial conformational changes at the B chain N terminal residues. These changes consist of a switch from an extended structure,

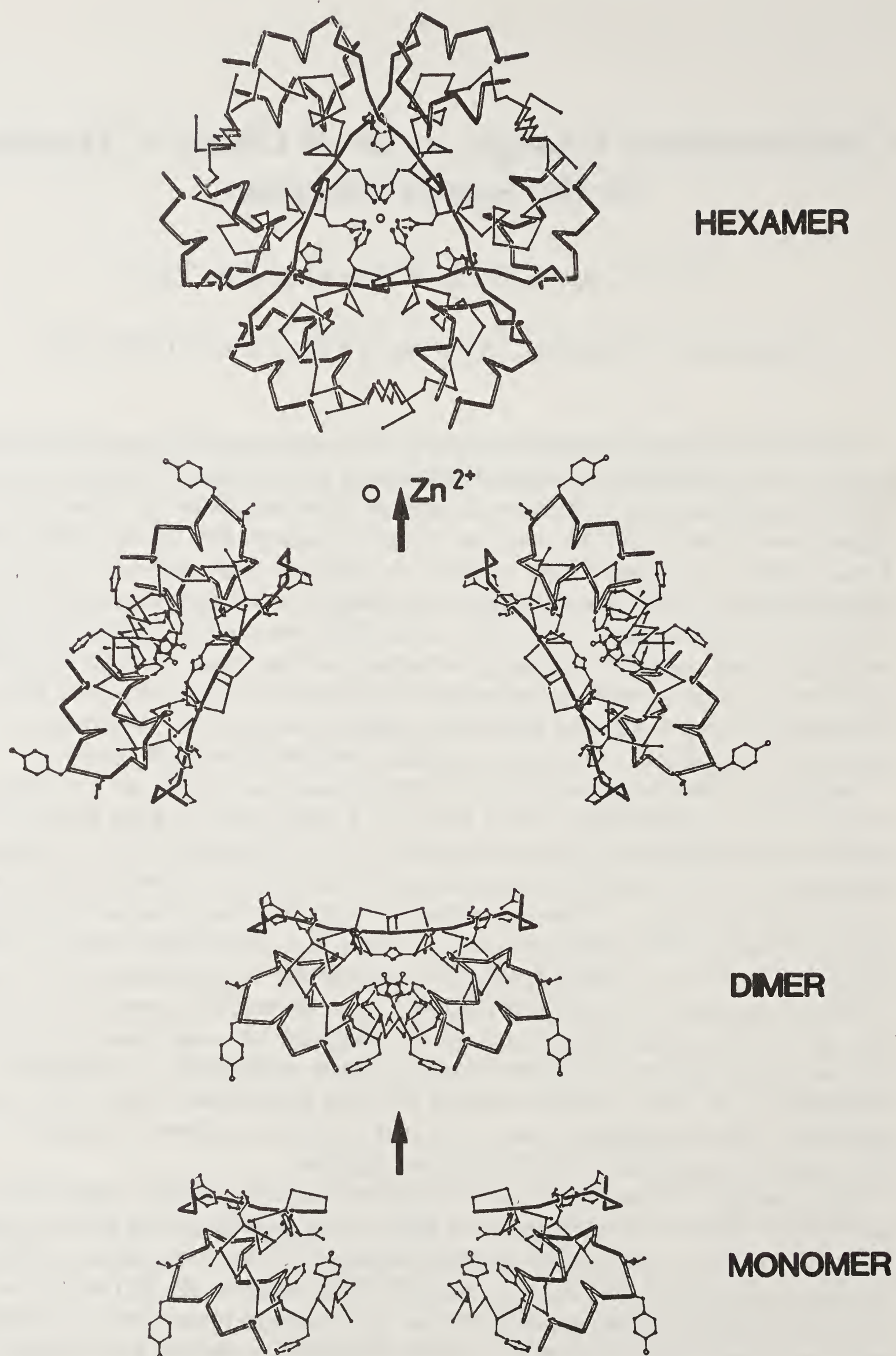


Figure 1. The assembly of the 2Zn insulin hexamer from monomers and dimers. The view is in the direction of the central 3-fold axis of the hexamer. In this figure only selected side chains are shown.

characteristic of the β sheet, to an α -helical organisation (4). The appearance of helix at B1–B8 creates a larger helix extending from B1–B20 (see figure 3 below). The transition requires a 20–30 Å movement at B1, as well as a 180° change in the dihedral angle of the disulphide bond at A7–B7. Since the B1–B8 residues lie between the dimers, changes in that structure affect the contacts and interactions made by the dimers in the hexamer.

The organisation of the different insulin hexamers and the conformations at B1–B8 are illustrated in figures 2 and 3.

The extended conformation at B1–B8 is seen in all six molecules of the 2Zn insulin hexamer. The crystals were grown from a solution of low ionic strength. The hexamer

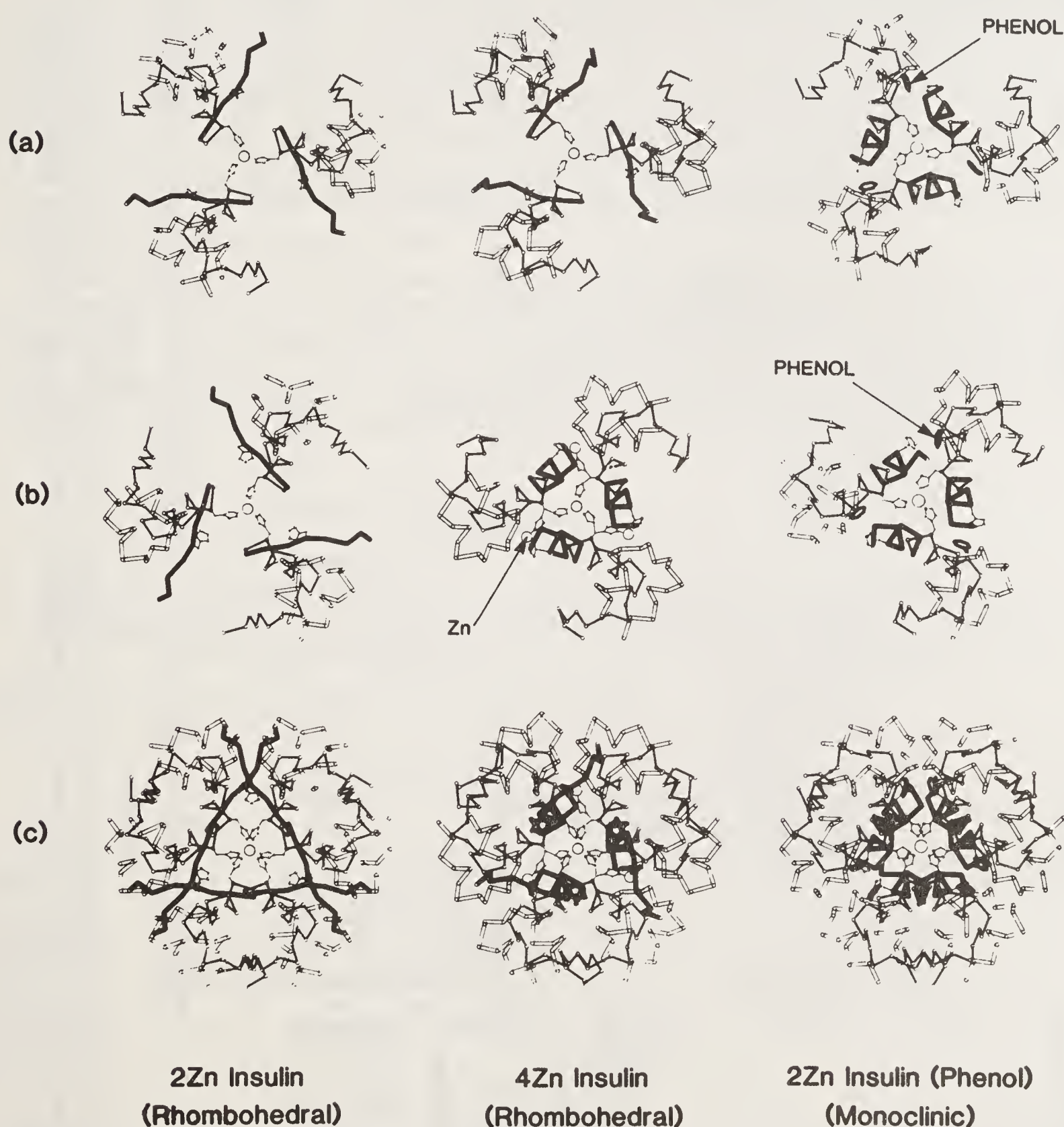


Figure 2. The organisation of the insulin hexamers from the 2Zn rhombohedral crystal (T6), the 4Zn insulin rhombohedral crystal (T3/R3) and the monoclinic phenol insulin crystal (R6). The alpha carbon atoms are shown together with the B5 and B10 histidine side chains, the zinc atoms and phenol molecules. The peptide segment B1–B8 is drawn with thick bonds, the rest of the B chain with thin bonds. The A chain is drawn with open bonds.

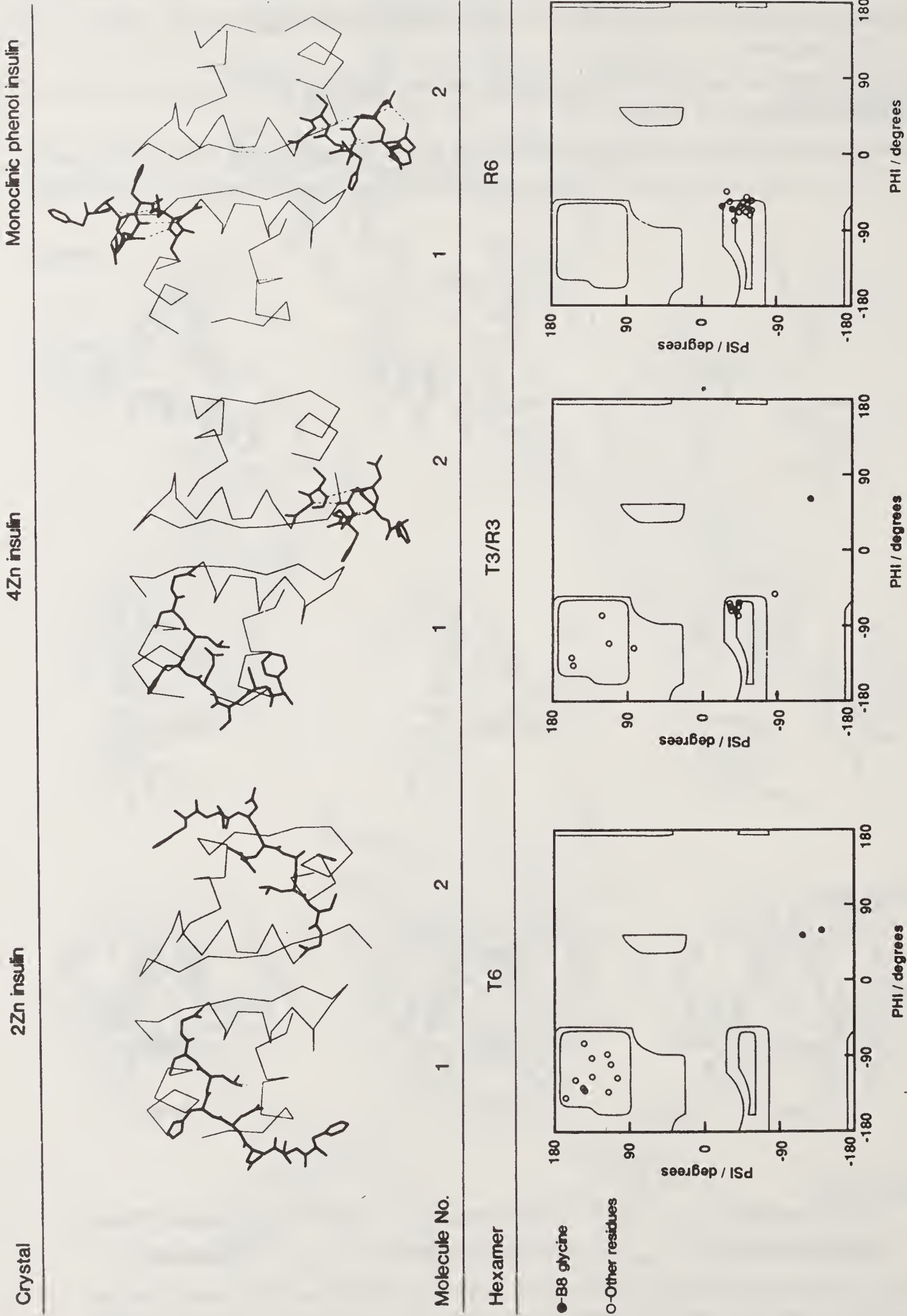


Figure 3. The structural changes at B1–B8 that occur in dimers in the T6, T3/R3 and R6 hexamers. The dimers are viewed in the direction of the local two-fold axis. Below each dimer is a Ramachandran plot showing the main chain dihedral angles of the residues B1–B13 from each molecule.

structure in this crystal is illustrated in figures 1 and 2(a). Insulin crystallised in the presence of 0.3 M chloride however is found to be arranged in a hexamer where 3 symmetrically related molecules contain B1–B8 in an extended structure, while the other three monomers contain helices at B1–B8, as illustrated in figure 2(b). This hexamer now no longer possesses 32 symmetry. There are movements in the A chain associated with the arrival of the helix which will not be discussed here (4). In the presence of phenol, zinc insulin hexamers undergo a further alteration in which now all six molecules have helix at B1–B8. The 32 symmetry seen in the original 2Zn insulin is restored though the contacts between dimers are obviously now quite different. The structure of this hexamer and its B1–B8 residues is shown in figure 2(c).

A convenient notation for the hexamers has been suggested (5) which is based on the capacity of the hexamer structure to bind phenol or zinc when the B1–B8 segment is helical. This conformation is referred to as R – corresponding to the R or high ligand affinity state in allosteric systems (6). In a parallel way the conformation where B1–B8 is extended and which does not bind phenol or zinc, is referred to as the T state, which in the allosteric convention has low or no ligand affinity. Thus the 2Zn insulin hexamer can be described as T6, the 4Zn insulin hexamer as T3/R3 and the phenol binding hexamer as R6.

The existence of the helical and extended conformations reflects the property of conformational flexibility which is intrinsic to peptides. It is unclear however whether these conformational alternatives have any biological significance. In this paper the interactions made by the α -helix and sheet structures within the hexamers are compared revealing some of the factors responsible for their stabilisation.

Experimental

The crystallographic details of the zinc insulin hexamers are given in table 1. The comparison of the 3 insulin hexamers was carried out using a Silicon Graphics

Table 1. Crystallographic details of zinc insulin hexamers.

	Space group	Cell parameters (Å)	Contacts of asymmetric unit	Associated molecules	Resolution of X-ray data	R cryst	Ref.
2Zn Insulin T6	R3	a = b = 82.5 c = 34.0	Dimer	—	1.5 Å	0.154	12
4Zn Insulin T3/R3	R3	a = b = 80.7 c = 37.6	Dimer	Cl ⁻	1.5 Å	0.18	9
2Zn Insulin (Phenol) R6	P2 ₁	a = 61.36 b = 61.71 c = 47.95 β = 110.8	Hexamer	Phenol m-cresol	1.8 Å	0.22	10
2Zn Insulin B8 Gly → Ser B13 Glu → Gln T3/R3	R3	a = b = 80.8 c = 35.7	Dimer	—	2.2 Å	0.19	B. Xiao, unpublished

(R = Σ||F₀| - |F_c||/Σ|F₀|)

work station using the CHARMM/QUANTA system (7). In these studies the conformational angles of the peptides from B1–B9 were reviewed and the environment of each residue was systematically investigated in order to identify van der Waals contacts. A contact was defined as a hydrogen bond (H bond) if the distance between the donor and acceptor atoms was less than 3.5 Å and the angle at the donor and acceptor was greater than 90° (8). Some estimates were made of the degree of hydration but these were limited in the monoclinic crystal (R6) and to a lesser extent in the 4Zn insulin crystal (T3/R3) by there being as yet an incomplete analysis of their water structure.

Atomic positions and thermal parameters have been refined by fast Fourier least squares minimisation calculations using the procedures of Konnert and Hendrickson to apply peptide bond and angle restraints (see table 1) (11). In these calculations the rms discrepancy between the reported values for the bond lengths and bond angles was *ca.* 0.02 Å and 5° respectively. The precision of the coordinates of the atoms ranges from about 0.1 Å to 0.2 Å for well defined atoms with thermal parameters (B) less than 15–20 Å². Atoms with high thermal parameters such as those in side chains and water molecules not anchored by H bond and other contacts have estimated errors of 0.3 Å or even higher.

Results

Conformational Angles at B1–B8

The dihedral angles of the peptides between B1–B13 are illustrated in figure 3 for the T6, T3/R3, R6 hexamers. The Ramachandran diagram reveals the change in the dihedral angles of the B1–B8 peptides in moving from the extended conformation associated with the T structure to the helical conformation associated with the R structure. The dihedral angles of B8 are noteworthy. This residue is invariant and occurs as glycine in all known insulin sequences. Its dihedral angles in the T structure correspond to a right-handed amino acid which in proteins of course can only be assumed by glycine. On converting to helix, the conformational angles of B8 glycine become typical of a left-handed amino acid.

Interactions in the B1–B8 Peptide

The major difference in the interactions made by the B1–B8 segment when in the extended or helical conformation is the main chain H bonding pattern. Details of these interactions are given in table 2. In the extended (T) conformation [illustrated in figure 3(a)], only the B4 carbonyl O and amide N and the B6 carbonyl O form β sheet-type H bonds. There are β bend or 3_{10} H bonds made by the carbonyl O's at B7, B8 and B9 to the N terminal turn of the B9–B20 helix. Since the extended B1–B8 peptide is mostly in a single strand the majority of its main chain polar atoms are hydrated. There is no H bond at B6 carbonyl O of molecule 2 in the crystal owing to the close packing of the A chain N terminus. In solution it is likely that both B6 carbonyl O's would be solvated. Out of the 9 amide NH groups in B1–B9, 7 are solvated and out of the 9 carbonyl O's, 5 are hydrated in the crystal.

In the helical (R) conformation, illustrated in figure 2(c), the main chain carbonyl O and amide N are all involved in the usual H bonding arrangement. In addition

Table 2. Interactions between main chain carbonyl (CO) and amide (NH) of B1–B8 in molecules 1 and 2 of 2Zn insulin (T6), 4Zn insulin (T3/R3) and monoclinic phenol insulin (R6). The water molecules are named as OW*n*, residue identifier where the residue identifier approximates to the distance along the C axis.

T6			T3/R3			R6							
Molecule 1			Molecule 2			Molecule 1			Molecule 2				
B1 NH ₃ ⁺	O ^ε A17	3·0	O ^ε A17	2·8	B1 NH ₃ ⁺	OB18	3·1	N ^ε 2 B5	3·0	B1 NH ₃ ⁺	O ^δ 1 B3	3·1	
			OW1, 62	2·7				OW4, 99	2·4		OW8, 55	2·3	
			OW1, 126	2·9									
			OW1, 68	3·0									
CO	OW1, 61	3·4	OW5, 14	2·4	CO	OW1, 126	3·2	125, OW1	2·8	CO	NH B5	3·1	
	OW5, 53	2·8	OW5, 16	2·8		OW1, 123	3·1	vdw A9			NH B4	3·3	
	OW4, 57	3·0	OW5, 20	3·4				vdw A1					
			OW5, 15	3·1									
B2 NH	OW1, 9	2·7	OW2, 62	2·7	NH	OB17	2·9	vdw A7		NH	vdw C ^γ 1 B2	O A8	3·0
											O ^δ 1 B18	3·2	
CO	OW2, 62	3·1	N ^ε 2 B4	2·3	CO	OW2, 54	3·1	NH B6	2·7	CO	NH B6	2·5	3·0
	OW2, 58	2·8	OW1, 9	3·0				NH B5	2·7		NH B5	3·4	3·3
	OW2, 60	3·1	OW1, 13	2·8									
B3 NH	OW7, 48	2·9	OW2, 23	3·2	NH	O ^ε 1 B21	3·2	No close contacts		NH	vdw C ^δ 1 B2	O ^δ 1 B18	3·2
CO	OW9, 43	3·4	OW5, 21	3·2	CO	OW1, 97	3·2	NH B7		CO	NH B6	3·3	3·1
	OW3, 44	2·5	OW1, 27	3·5							NH B7	2·9	
	CO A11	3·1	CO A11	2·7	NH			OW4, 100	2·7	NH	O B2	3·2	3·4
CO											O B2	3·3	
	NH A11	3·2	NH A11	3·3	CO	NH B6	3·4	NH B8	2·8	CO	NH B8	3·1	3·0
	β		β			OW4, 150	2·6	α					
B5 NH	OW7, 47	2·5	OW2, 22	3·0	NH	OW1, 143		OB2	2·7	NH	OB1	3·1	3·4
	(D)							OB3	3·0		OB2	2·9	3·2

(Continued)

Table 2. (Continued)

T6				T3/R3				R6							
Molecule 1				Molecule 2				Molecule 1				Molecule 2			
CO	OW3,47	2.9	OW1,21	3.2	CO	OW1,54	3.1	HO B16	2.7	CO	OH B16	2.4	NH B8	3.3	
	OW2,53	3.2	OW2,25	3.0		O ^{e1} ,A4	3.4	NH B9	3.2		NH B8	3.4	NH B9	3.1	
	OW1,51	3.2	OW6,19	2.6							NH B9	3.4	HO B16	2.5	
			OW4,20	3.5											
B6 NH	OA6	2.8	OA6	2.9	NH	OA6	2.8	OB2	2.7	NH	OB2	2.5	OB2	3.0	
	OB11	3.4						OB3	3.3		OB3	3.3			
CO			OW2,20	2.8	CO	O ^{e1} ,A4	3.4	NB9	3.2	CO	NH B9	3.0	NH B10	3.3	
						OW1,54	3.1		2.7		NH B10	3.5	NH B9	3.1	
B7 NH	OW1,51	3.0	OW1,21	2.8	NH	OW1,54	3.4	OB3	3.0	NH	OB3	2.9	OB3	3.1	
CO	NH A10	3.1	NH A10	3.1	CO	NH B11	3.4	NH B11	3.0	CO	NH B11	2.9	NH B11	2.9	
	OW4,48	3.5				NH B10	3.2	NH B10	3.2		NH B10	3.0	NH B10	3.1	
											NH B9	2.9			
B8 NH	OW2,49	3.0	OW1,26	3.0	NH	OW7,53	2.9	OB4	2.8	NH	OB4	3.1	OB4	3.0	
	OW4,51	3.1	OW1,17	3.4							OB5	3.4			
CO			vdw												
	NH B11	3.1	NH B11	3.1	CO	NH B12	3.2	NH B11	3.2	CO	NH B11	3.4	NH B12	2.8	
	NH B10	3.2	NH B10	3.1		NH B11	3.2	NH B12	2.9		NH B12	3.0	NH B11	3.0	
	NH B12	3.4	OW1,17	2.8											
B9 NH	OW4,51	3.0													
	OW3,51	3.1	OW1,16	3.0	NH	O ^r B9	2.8	OB5	3.2	NH	OB6	3.0	OB6	3.1	
CO						OW1,59	3.5	OB6	3.2		OB5	3.4	OB5	3.1	
								OH B16	3.3						
	NH B13	3.1	NH B13	3.1	CO	NH B13	2.9	NH B13	3.4	CO	NH B13	3.4	NH B12	3.3	
						NH B12	3.5	O ^{e2} B13	2.6		OW6,47	3.2			

the B16 Tyr OH makes a strong H bond to the carbonyl O of B5 His. This, and the helical H bond B5 CO–B9NH satisfy the carbonyl's full H bonding potential. Note there is a long H bond (*ca.* 3.5 Å) between B4 carbonyl O and B13 NH which also occurs in all the other helices present in the R structure.

Sidechain Interactions

Some of the side chains in the extended and the helical structures make strong interactions which are probably important in defining the B1–B8 conformation. These are described briefly in table 3.

Table 3. The B1–B8 side chain contacts in the extended (T) and helical (R) conformations.

T (as in 2Zn insulin)		R (as in monoclinic insulin)	
B1 Phe	Buried at the hexamer edge, its α -amino group makes a long salt bridge to the A17 Glu of the adjacent dimer. The aromatic ring is inserted into a non-polar pocket in the adjacent dimer from which it makes weak van der Waals contact to its two-fold equivalent.	B1 Phe	The phenylalanine side chain is directed into solvent and makes contact mainly with water. (There are lattice contacts with an adjacent hexamer in the crystal.)
B2 Val	The valine side chain is partially buried, it makes nonbonded contacts to solvent and to protein.	B2 Val	There are no close protein contacts, the valine is mostly solvated. (The adjacent hexamer in the crystal makes some contacts.)
B3 Asn	This side chain is on the hexamer surface and is extensively hydrated.	B3 Asn	An external polar side chain, surrounded by solvent. (There are some contacts in the crystal with symmetry related hexamers.)
B4 Gln	This side chain is on the surface between dimers in the hexamer. It is directed towards the interior from the extended peptide. The side chains in both molecules make four contacts to protein and are extensively hydrated. In molecule 2 the side chain is disordered.	B4 Gln	The side chain is seen sometimes to H bond to B3 O ^{δ1} , the adjacent residue. Otherwise it is mostly hydrated.
B5 His	The imidazole N ^{ϵ2} H bonds to A9 O in molecule 1 of the 2Zn insulin crystal. This interaction probably persists in solution. It is seen in dimeric and monomeric insulin crystals. Lattice contacts in 2Zn insulin affect the H bonding of N ^{δ2} (molecule 1) and the B5 imidazole in molecule 2.	B5 His	The imidazole group is at van der Waals distance to the phenol. In the T3/R3 hexamer, the B5 His in the helical arrangement is coordinated to the off axial zinc ions. When this site is not occupied by zinc, it is occupied by a water molecule which H bonds to the imidazole ring.
B6 Leu	The side chains are pointed down into the dimer–dimer interface where they pack into non-polar surfaces but are not completely buried.	B6 Leu	The non-polar side chain makes van der Waals contacts to the phenol in the R6 hexamer. In the T3/R3 system, this side chain,

(Continued)

Table 3. (Continued)

T (as in 2Zn insulin)		R (as in monoclinic insulin)	
		when in the helical arrangement, is partially hydrated with contacts to B10 His. These Leu side chains are arranged together above the axial zinc site, reducing Zn exchange.	
B7 Cys	The dihedral angles show the A7–B7 bond to be right handed. The S–S bridge is exposed and in contact only with water.	Bt Cys	The dihedral angles of the A7–B7 disulphide bond are left handed. They are largely occluded from solvent by the packing of the B1–B8 helices
B8 Gly	This residue takes a conformation with $\psi = -20^\circ$ and -19° and $\phi = -62^\circ$ and -65° in molecules 1 and 2 respectively. This residue has a right-handed conformation which allows the CO to H bond to amide NH of B13, with a weaker interaction to B12 NH. There are close approaches to C $^\alpha$ from B16 tyrosyl OH (3.4 and 3.3 Å) and longer from B26 tyrosyl OH (6.1 and 5.6 Å) in molecules 1 and 2 respectively.	B8 Gly	The residues are in a helical conformation. The B8 C $^\alpha$ is approached by both B26 Tyr OH from the same monomer and B16 Tyr OH from the other molecule in the dimer.
B9 Ser	The side chain O $^\gamma$ is directed into the polar central channel where it forms H bonds to water molecules.	B9 Ser	The O $^\gamma$ atom can make H bond contact to B13 O $^\epsilon$; it sometimes hydrates without contact to protein.

Interactions at B8 Gly

Inspection of the interactions at B8 in the T structure shows that the carbonyl O can make a satisfactory H bond to a main chain NH only with a right-handed conformation. The H bonding made in this conformation by B8 CO is almost ideal – two H bonds are made to B10 NH and B12 NH at the N terminus of the helix between B9–B20. This satisfies both these N terminal peptide nitrogens otherwise unable to make a peptide H bond, as well as satisfying the CO H bonding potential completely. The same H bonding contacts are made in the R state when the B8 Gly is incorporated in the helix structure.

The B8 NH is H bonded to water in the T or extended conformation. In the helical conformation it H bonds to B4 carbonyl O. There are in the R structure however very close contacts between B8 C $^\alpha$ and the OH of B16 Tyr which is then able to be H bonded to B5 carbonyl O.

Interactions of B8 Serine in B8 Gly \rightarrow Ser (B13 Glu \rightarrow Gln)

The insulin containing the double mutant B8 Gly \rightarrow Ser and B13 Glu \rightarrow Gln (hereafter referred to as B8 Ser) was crystallised in low salt conditions where normally the hexamer would take the T6 organisation, i.e. the B1–B8 residues will have an extended

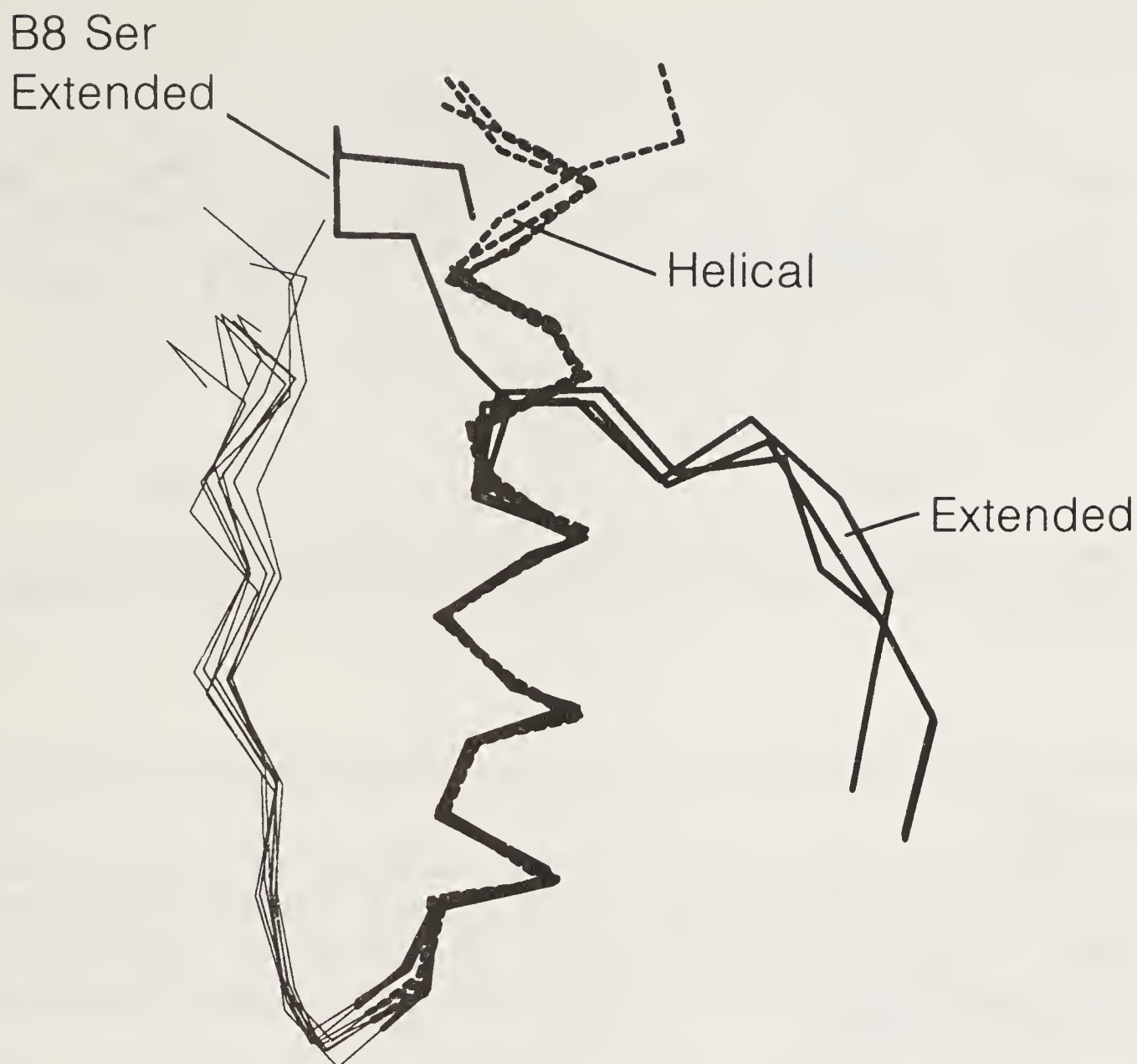


Figure 4. C^α trace of molecules 1 and 2 of 2Zn insulin, 4Zn insulin, monoclinic insulin and B8 Ser insulin overlapped by least squares of the main chain atoms of the central helix B9–B19. Residues B1–B20 of molecule 1 are drawn with solid bold lines; residues B1–B20 of molecule 2 are drawn with dashed bold lines.

conformation. The mutant insulin crystals were rhombohedral, with two molecules in the asymmetric unit and with the hexamer constituting the unit cell (see table 1). But the structure does not adopt a T6 conformation. In spite of the absence of chloride or phenol the hexamer has a T3/R3 character normally associated with high chloride concentration.

The extended structure in molecule 1 of the B8 Ser mutant insulin is markedly different in conformation to that seen in the native T3/R3 and T6 hexamers. Figure 4 shows the overlap of C^α traces of the B chain of both molecules of 2Zn insulin, 4Zn insulin, monoclinic insulin and B8 Ser insulin. In this T3/R3 hexamer organisation the B8 serine OH hydrogen bonds to water and its carbonyl O hydrogen bonds to B11 NH but not to B12 NH. A more detailed picture of the B1–B8 conformation in this molecule is shown in figure 5(a).

The conformation of B1–B8 in molecule 2 of the T3/R3 hexamer of the mutant insulin is helical [see figure 5(b)] and similar to that seen in the native 4Zn insulin T3/R3 crystal. There are inevitable changes in the contacts at B8 however. An H bond exists between B8 Ser OH and the carbonyl O at B4 replacing the interaction with the B8 peptide NH. The B16 Tyr side chain is displaced by the bulk of the serine

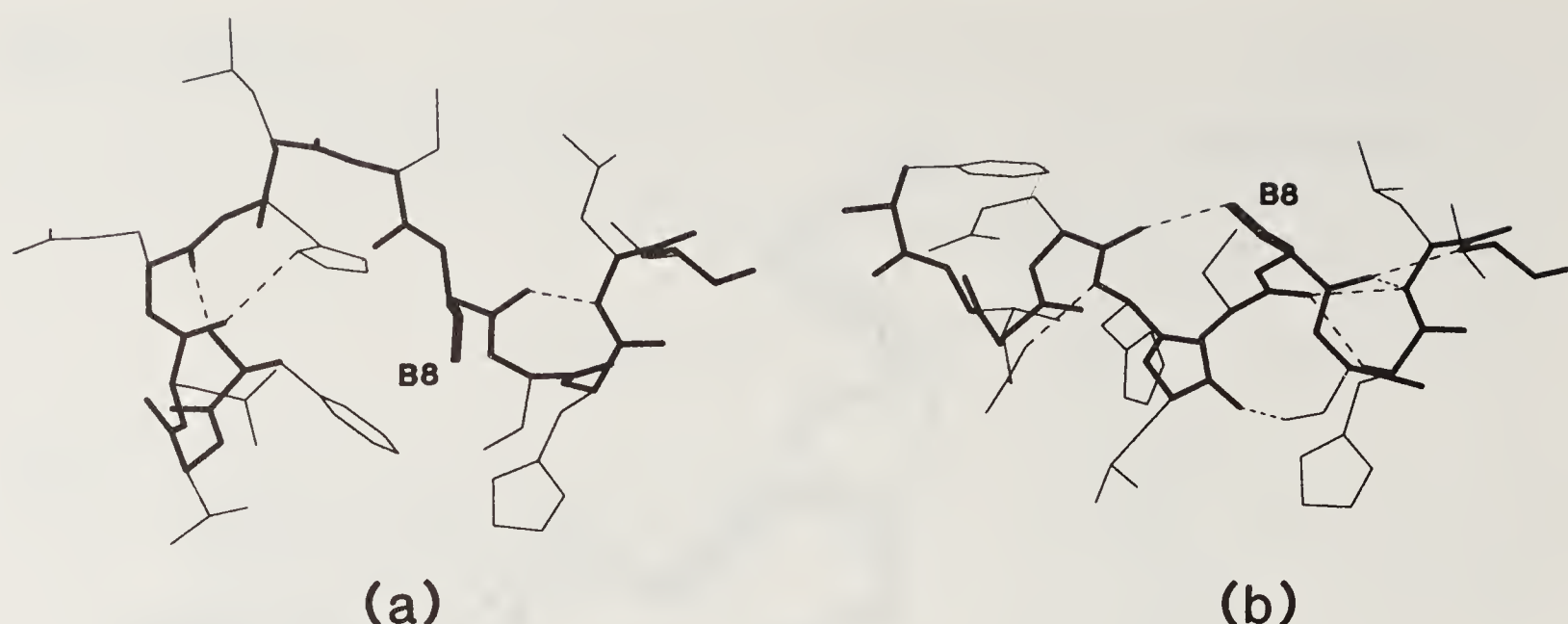


Figure 5. Conformation and internal hydrogen bonding of residues B1–B12 of (a) molecule 1 and (b) molecule 2 of B8 Ser insulin. Main chain atoms are connected by bold lines.

side chain at B8 and its OH group H bonds to the B5 imidazole N^δ instead of the B5 carbonyl O.

As a consequence of these sequence changes at B8 and B13 the B9 ser OH and carbonyl O form H bonds from the R helical structure directly to the mutated B13 Gln amide. A similar interaction is seen in the R structures of 4Zn insulin and the monoclinic phenol insulin. Here however the B13 carboxylic acid H bonds only to the serine OH (see table 3).

Discussion

The B1–B8 Structure and the Insulin Hexamer

The B1–B8 segments in the 2Zn insulin (T6) hexamer appear to link the dimers together (12). Their N terminal residues are buried in a non-polar pocket of the adjacent dimer. The two segments arch out from the central B chain helix in opposing structures and enclose between them the interface between the dimers (see figure 2). Thus there are very few contacts between the B1–B8 peptides in the T structure; only the equivalent phenylalanine rings of the symmetry related B1 Phe make weak van der Waals contacts. The two B1–B8 strands are quite closely related by the two-fold symmetry axis of the dimers. This organisation suggests the N terminal sequence at the B chain is structurally important in the design of the T6 insulin hexamer. The conformations and positions of the B1–B8 peptides are defined in the T6 hexamer by 3 β sheet type H bonds to B4 and 11 (CO and NH) and between B6 and A6 (NH) [see figure 6(a)]. This arrangement obeys the local two-fold symmetry of the hexamer well.

In the T3/R3 or 4Zn insulin hexamer one B1–B8 segment remains in the extended conformation while the other is α -helical. This disappearance of extended conformation removes the contacts between the N terminal B1 phenylalanyl side chains [see figure 6(b)]. Consequently the B1–B8 residues in the extended (T) conformation collapse into the volume between the dimers, previously occupied by B2 valine. As illustrated

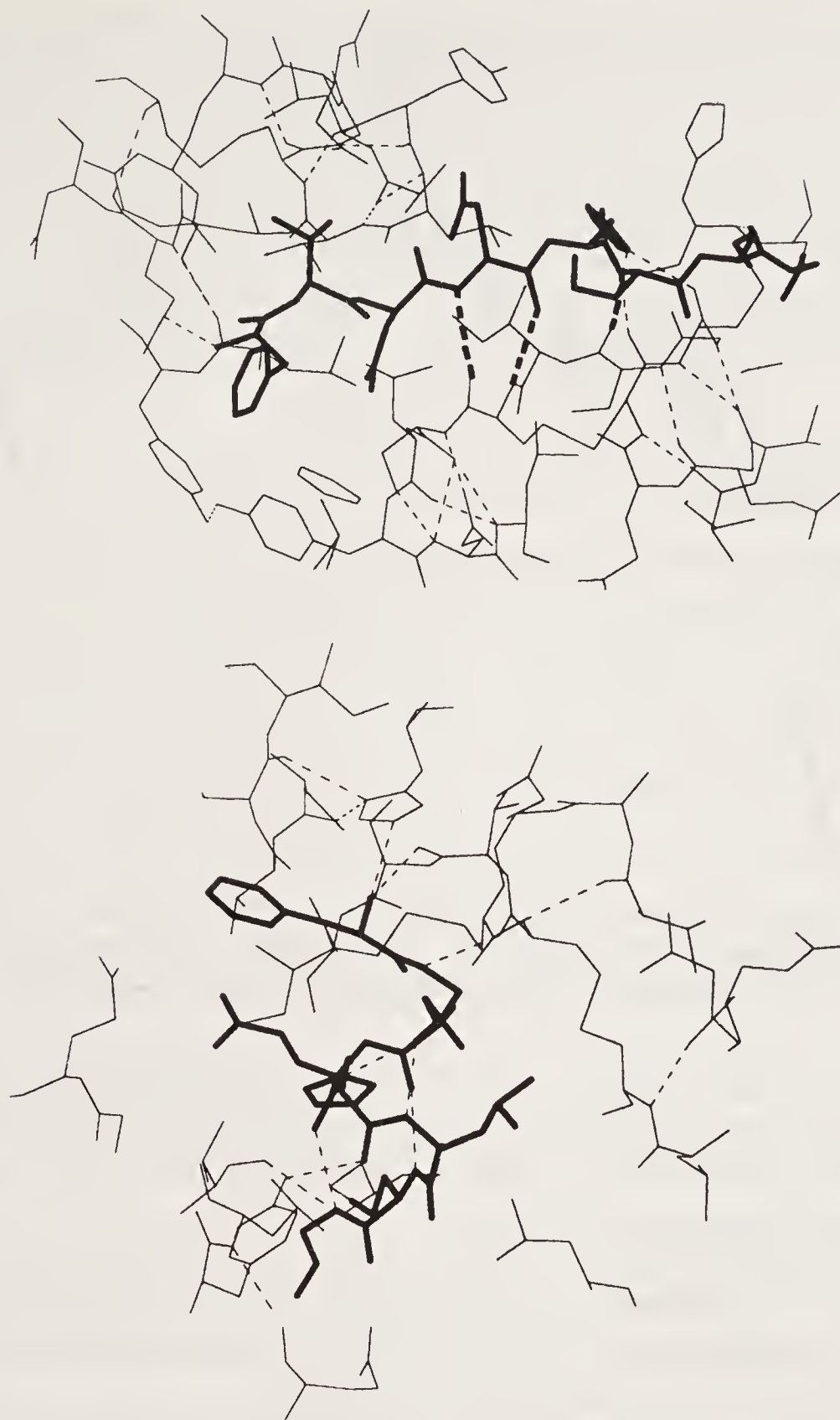


Figure 6. Interactions of B1–B8 in the extended (T) and helical (R) structures. (a) Residues B1–B8 of molecule 1 2Zn insulin are shown in bold lines. The 3 important β sheet type hydrogen bonds to the A chain are shown in thick dashed lines. (b) Residues B1–B8 of molecule 2 4Zn insulin are shown in bold lines.

in figure 7, the position of B1–B4 has therefore particularly changed, the H bond contacts with the A chain are lost and replaced by contact with water molecules.

The T3/R3 symmetry as a result is now reduced to three-fold symmetry and the local two-fold symmetry is obeyed only partially within each dimer. Moreover, the dimers move in response to the altered B1–B8 structures and their local two-fold axes, now only approximate, no longer intersect the central 3-fold axis.

The three helical extensions at B1–B8 form a pyramid around the axis on the surface but do not make any close contacts. The three B10 His imidazole rings are at van der Waals separation in their coordination structure around the zinc on the 3-fold axis. Above the histidines the three B6 leucines pack together at 4–5 Å

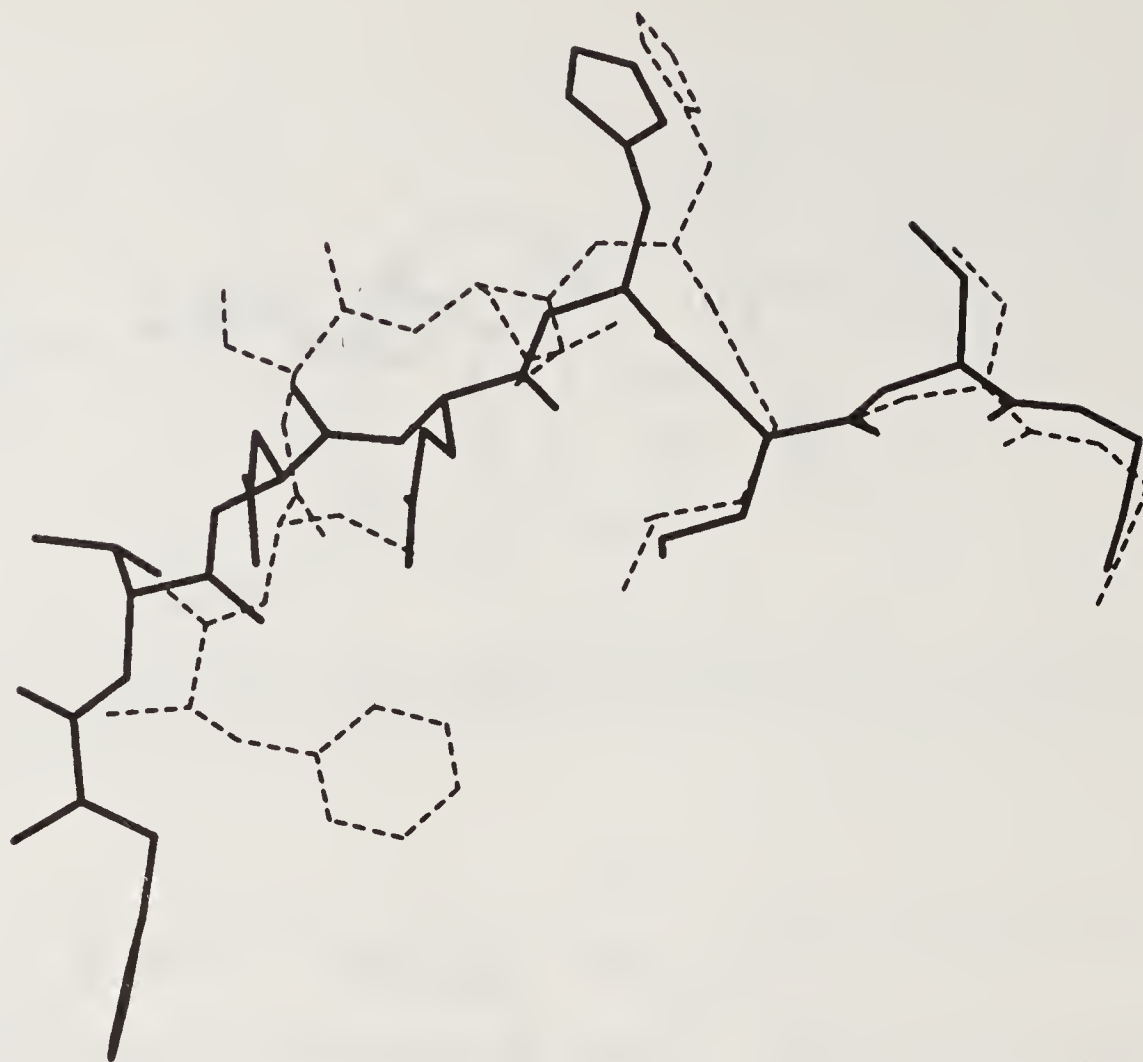


Figure 7. Change in extended conformation of B1–B8 between T6 and T3/R3 structures. Residues B1–B8 of molecule 1 2Zn insulin are shown in bold lines. Residues B1–B8 of molecule 1 4Zn insulin are drawn with dashed lines.

separation, close enough to prevent water molecules populating the cavity in the axis but further than the van der Waals distance. Finally, the loss of extended peptide at B4 and B6 leads to the polar atom A6 CO and NH and A11 CO being H bonded instead to well-defined water molecules.

The most important interaction made by the helices in the T3/R3 hexamer is through the B5 His which can coordinate zinc ions in an off axial site (see table 3). Because the zinc coordination also involves B10 His from an adjacent helix this links the helices (and therefore the 3 T/R dimers) together.

The formation of the T3/R3 or the 4Zn insulin structure depends on the chloride ion concentration which should be greater than *ca.* 0.3 M. The analysis of the crystal structure shows that chloride ions bind to the on-axial tetrahedral zincs associated with the R structure insulin. Chloride ions are also involved in the off-axial tetrahedral zinc site. Thus although the chemical energy for stabilising the hexamer (and the T and R states) comes in part from zinc coordination, the conformation at B1–B8 in the R structure is clearly governed by the effects chloride has on stabilising the protein conformation (13).

In the R6 hexamer formed with phenol the dimers are symmetrical and this local axis of two-fold symmetry intersects the central 3-fold axis as can be seen by inspection of figure 3. Thus the 32 symmetry seen in the T6 hexamer is restored. The helices pack together in the same rather loose way as the equivalent structure in the T3/R3 hexamer. There is no off-axial zinc site in the R6 structure; instead a phenol molecule binds at this site interacting with B5 imidazole (van der Waals) in one dimer and with A6 and A11 (CO and NH respectively) from the other dimer. These H bonds from the phenol replace in part those lost by the removal of the β sheet interactions

in the T6 hexamer between B4 and A11 and B6 and A6. The phenol molecule in addition partially fills a cavity between the dimers and makes a series of loose van der Waals contacts between the dimers.

The 3 helical structures have a further effect on the hexamer. Together they form a trigonal pyramid and although the side chains are not closely packed they greatly reduce access to and escape from the zinc site. This effect has been observed (reference 5 and M. Dunn, private communication). This explains how the presence of the helix at B1–B8 through the interactions with phenol (or zinc and chloride) leads to increased stability of the hexamer (14).

The Conformation at B8 in the B8 Gly → Ser Mutant

The glycine at B8 is invariant throughout native insulin sequences. The X-ray analysis of the mutant insulin B8 Gly → Ser shows that the presence of the side chain at this residue dictates a significantly different conformation at B1–B8. Crystals of B8 Gly → Ser were grown in conditions that with the native hormone led to the T6 structure in which B8 Gly assumes right-handed amino acid dihedral angles. The crystals were rhombohedral but had axial lengths more characteristic of the T3/R3 (4Zn insulin) crystal structure (see table 1). In the 4Zn insulin crystal one molecule has an extended structure at B1–B8 with a right-handed conformation at B8 while the other has B1–B8 in helix.

Inspection of the electron density after refinement reveals that in molecule 1 of the asymmetric unit the B1–B8 is extended but has distinctly different conformations and makes quite different contacts to those seen in its equivalence in 4Zn insulin. It does not have a β sheet organisation or H bond to the A chain and is incorporated in the helix with dihedral angles corresponding to a left-handed amino acid.

The structure of B1–B8 of molecule 1 of B8 Ser is shown in figure 5(a). The most striking feature is the H bonds formed by the B1 α amino group. One of these is to B4 carbonyl O and the other to B5 carbonyl O, again putting a conformational restraint on the peptide chain. The dihedral angles at B8 Ser in this conformation are $\psi = 138^\circ$, $\phi = -169^\circ$, consistent with a helix or turn. It can be seen that in this B8 Ser structure B1–B8 continues to turn in contrast to the native molecule. This turn is presumably stabilised by the B8 Ser O γ hydrogen bond. The Tyr B16 side chain is displaced by the extra bulk of the serine side chain.

In the second molecule of the asymmetric unit the B1–B8 residues form an α helix, illustrated in figure 5(b). It is very similar in structure and contacts to that which occurs in the native 4Zn insulin crystal. There are however different contacts at B8 Ser. The side chain B8 serine O γ is H bonded to the carbonyl O of B4 Gln, and the B9 serine O γ is H bonded to the carbonyl O of B5 His, as described earlier. The serine side chain at B8 displaces B16 and B26 tyrosines. Now the B16 tyr OH is H bonded to the B5 His N δ rather than the B5 carbonyl O.

Conclusions

The striking variation in the conformations at B1–B8 in the insulin hexamer depends on the flexibility of the protein and the capacity of the side chains to make suitable interactions in each of the two conformations. It appears that the crucial phenomenon

is the contact made by B5 His when in the R (helical) conformation. In the T3/R3 structure it binds zinc (or water) and through this forms a link to the adjacent dimer. In the R6 structure, the phenol molecule brings about a similar bridging effect through van der Waals contacts with the imidazole ring and H bond and other close contacts with residues in the adjacent dimer.

Another key residue for the existence of the two conformations is the B8 Gly invariant in all known insulin sequences. In the T6 hexamer it exhibits the conformations of a right-handed amino acid. The absence of a side chain first allows the glycine to make good main chain H bond contacts, and second, it permits the close intramolecular packing of the tyrosine B16 OH group and the approach ($\sim 5 \text{ \AA}$) of the B26 tyrosyl OH of the other molecule in the dimer. In the helical conformation there prove to be structural advantages for glycine at B8. First, as in the T structure, it allows intramolecular close packing of B16 tyrosine which favours dimer formation. Second, in this arrangement B16 Tyr OH forms a good H bond to the CO of B5 His in the helix thus completing its full H bond potential.

The factors that govern the existence of the extended or helical conformations are becoming better defined. It is clear that the energy difference between the two conformations is small and that the two states are readily interconvertible and in solution will be in rapid equilibrium. Thus it should be possible to manipulate the conformational behaviour at B1–B8 in hexamers with molecules other than phenol or chloride. The techniques of protein engineering are allowing us to explore with new precision and power the relationships between the sequence at B1–B8 and its conformational properties.

Acknowledgements

We thank Novo-Nordisk A/S for financial support and supply of native and mutant insulin material. This work was also supported by the Science and Engineering Research Council, the Medical Research Council and the Juvenile Diabetes Foundation. We gratefully acknowledge discussions and assistance provided by Dr. Jens Brange and Mrs. Eleanor Dodson.

References

1. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963) *J. Mol. Biol.* **7**, 95–99.
2. Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimmer, B. & Sheat, S. (1969) *Nature (London)* **224**, 491.
3. Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Reynolds, C. D. & Vallely, D. (1981) in *Structural studies on molecules of biological interest*, eds. Dodson, C. G., Glusker, J. P. & Sayre, D. (Oxford University Press) pp. 527–546.
4. Bentley, G. A., Dodson, E. J., Dodson, G. G., Hodgkin, D. C. & Mercola, D. A. (1976) *Nature (London)* **261**, 166–168.
5. Kaarsholm, N. C., Ko, H.-C. & Dunn, M. J. (1989) *Biochemistry* **28**, 4427.
6. Monod, J., Wyman, T. & Changeux, J. P. (1965) *J. Mol. Biol.* **12**, 88.
7. QUANTA/CHARMm software provided by Polygen Corp.
8. Hohne, G. & Kretschnner, G. (1982) *Studia Biophysica* **87**, 23–28. (see also reference 12 below).
9. Smith, G. D., Swenson, D. C., Dodson, E. J., Dodson, G. G. & Reynolds, C. D. (1984) *Proc.*

- Natl. Acad. Sci. U.S.A.* **81**, 7093–7097. Also Xiao, B., Derewenda, U. and Dodson, E. J. (unpublished results).
10. Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Reynolds, C. D., Smith, G. D., Sparks, C. & Swenson, D. (1989) *Nature (London)* **338**, 594–596.
 11. Konnert, J. H. & Hendrickson, W. A. (1980) in *Computing in Crystallography*, eds. Diamond, R., Ramaseshan, S. & Venkatesan, K. (Indian Acad. Sci., Bangalore) pp. 1301–1323.
 12. Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N. & Vijayan, M. (1988) *Philos. Trans. R. Soc. London* **B319**, 369–456.
 13. Schlichtkrull, J., (1958) PhD Thesis, University of Copenhagen.
 14. Brange, J., Skelbaek-Pedersen, B., Langkjaer, L., Damgaard, U., Ege, H., Havelund, S., Heding, L. G., Jorgensen, K. H., Lykkeberg, J., Markussen, J., Pingel, M. and Rasmussen, E. (1989) *Galenics of Insulin* (Munksgaard, Copenhagen).

Protein-Crystallography and Enzyme Catalysis

Georg E. Schulz

Institut für Organische Chemie und Biochemie der Universität,
Albertstr. 21, 7800-Freiburg i.Br., Germany

Undoubtedly, the understanding of enzyme catalysis has been revolutionized by the analyses of the spatial structures of these enzymes, which was started with hen egg white lysozyme and has been in progress for about 25 years with ever-increasing intensity. The methods to be followed for these analyses are quite diverse and greatly depend on the enzyme species. The analytical techniques have improved over the years so that present-day protein-crystallography differs appreciably from the methods used in early times.

Modern Protein-Crystallography

For producing crystals, proteins have to be isolated to high purity. Homogeneity is required not only with respect to the covalent structure but also with respect to the spatial structure which for the most part is defined by the polypeptide conformation. The situation is reflected by the fact that the first structures were obtained from the very stable and massively produced enzymes lysozyme, chymotrypsin, ribonuclease and carboxypeptidase. These enzymes have to be stable because they are secreted into the extracellular spaces where they have to withstand very unfavorable environments.

The more interesting enzymes are intracellular and occur in much smaller quantities. As sketched in figure 1, this means that from a kilogram of living material one may eventually obtain only about a milligram of pure enzyme. This has to suffice for crystallization and X-ray structure analysis on the one hand and also for the required sequence analysis at the peptide level on the other. The most serious problem is the purification of the protein by a factor of about a million, which requires meticulously designed multi-step procedures that are prone to destroy the chemical and/or conformational homogeneity of the molecule ensemble.

While chemical deterioration is mostly deamidation of Asn and Gln that can be usually detected by running an isoelectric focussing gel, no test of similar simplicity is available for detecting partially denatured, i.e. conformationally deteriorated proteins. Furthermore, one has to keep in mind that the detection of a problem is only the first step in solving it. Avoiding the problem is more important, but may be difficult in protein purification.

A substantial advance was the introduction of recombinant protein production. For this purpose, the gene of the interesting protein is isolated, incorporated in a bacterium and expressed in large quantities. Apart from obtaining larger masses of protein, it is crucial that proteins are produced as the major component of the bacterial cell so that

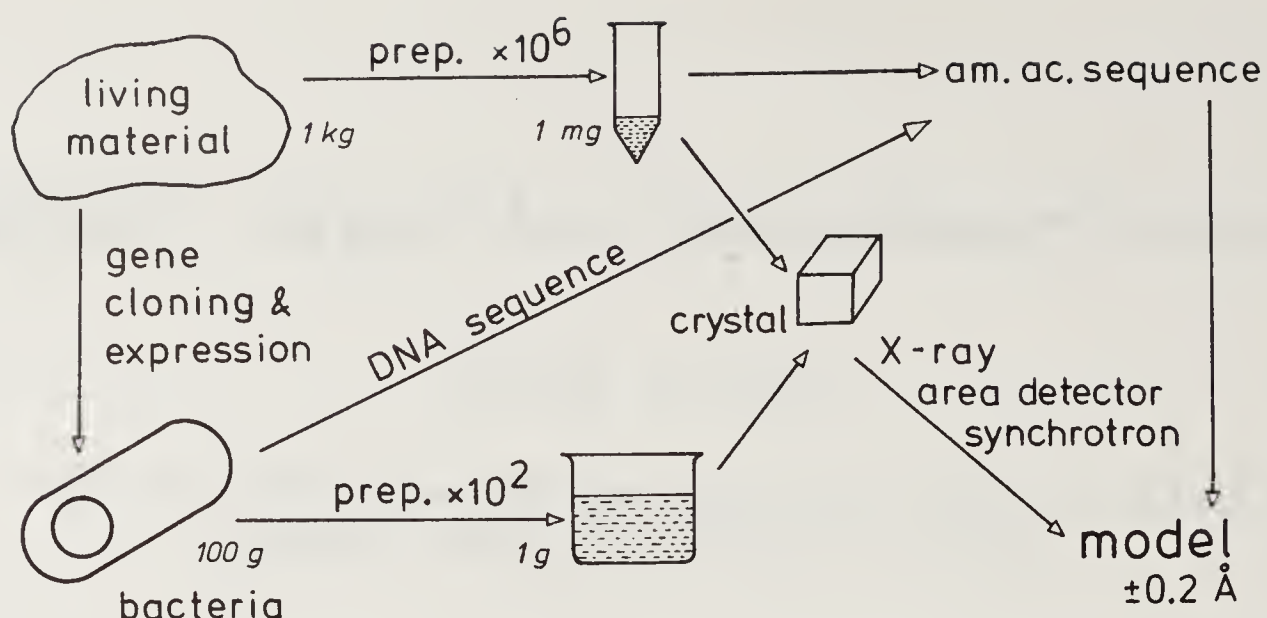


Figure 1. Present-day protein-crystallography. The analysis is largely facilitated by producing proteins in high quantities using efficient expression systems in plasmids of bacteria, DNA-sequence analysis as well as strong sources and fast detectors in X-ray diffraction analysis.

the purification factor no longer ranges around a million but around a hundred (figure 1). This shortens the isolation procedures to such an extent that the proteins no longer deteriorate during the preparation and are therefore easier to crystallize. Moreover, with much larger amounts of protein at hand, chances of crystallization improve appreciably, because many more trials can be made. In addition, the sequence analysis can now be pursued at the DNA-level, which is much simpler than at the peptide level. Taken together, the chances for a successful structure analysis have dramatically increased with the introduction of recombinant proteins.

Advances in the chemical procedures have been accompanied by important improvements in the X-ray diffraction machinery. Protein-crystallography started with X-ray films and more or less automated four-circle diffractometers. Now, these are only used for initial crystal screening. All major data collection is done using direct electronic films like multi-wire (1) or scintillation area detectors (2) or an indirect electronic film like the image-plate (3). The still tedious search for suitable heavy atom derivatives has become easier because

- (i) many problems are now being solved using molecular replacement methods based on a homologous structure that is already known,
- (ii) the number of necessary heavy atom derivatives has appreciably decreased due to the phase improvement methods available, in particular, the solvent flattening procedure (4), and
- (iii) the required resolution for the multiple isomorphous replacement phases from heavy atom derivatives has now reached the 3.0 to 3.5 Å range for all cases where native diffraction data up to around 2 Å resolution are available. Better experimental phases are no longer necessary, because the new structural refinement methods using simulated annealing (5) have a large radius of convergence and can therefore start from a rather ill-defined structure.

While the chemical advances have required cooperation by laboratories with different specialities, the physical advances can be adopted more easily, they are just expensive. The prices of area detectors and computers for simulated annealing refinements are about an order of magnitude higher than those of a four circle diffractometer and computers for conventional structure analyses and refinements.

Crystallizing Protein Conformations

For organic molecules with some conformational freedom it is well-known that the crystallizing species can represent quite a minority of the conformational ensemble assumed in solution. Therefore, organic chemists generally refrain from deriving conformational information from crystal structures. If this argument would apply to a protein, the conformation of which is crucial for all reactions, protein-crystallography would be useless.

In contrast to organic molecules, however, the native protein structure is quite uniquely defined among the myriad of conformations available for a polypeptide chain. This was already demonstrated with the early structure analysis of α -chymotrypsin (6). The α -chymotrypsin crystals contain two essentially identical enzyme molecules in the crystallographic asymmetric unit, that means in different crystal packing arrangements. One could therefore conclude that the observed conformation is the major species in solution, because a minor species that is stabilized in two different arrangements of crystal packing contacts would be most unlikely.

Meanwhile, many similar observations were made with other proteins, either with more than one molecule in the asymmetric unit or with the same molecule in different crystal forms. That protein crystals in general contain the major conformational species in solution was further supported by nuclear magnetic resonance measurements in solution yielding essentially the same structures as in the crystals (7).

Catalytic Activity in the Crystal

Even after a large number of observations demonstrated that crystalline and dissolved proteins are essentially identical it was still not clear whether crystalline enzymes are catalytically active, because there were differences in the side-chain conformations at the surface; and catalysis depends on side-chains. As a matter of fact, this question is difficult to answer, because any measured catalytic reaction in the crystal will be at least as slow as the substrate diffusion to the catalytic centers in the crystal, which takes several seconds. Moreover, it is most difficult to prove that a slow reaction of crystalline enzyme is not mimicked by a small amount of fast-reacting enzyme freed by partial crystal dissolution.

The situation becomes clearer if the catalytic cycle of an enzyme runs through a stable intermediate as with glutathione reductase. The reaction of this flavo-enzyme is



where the color of the stable intermediate EH_2 is brownish and of the standard oxidized state E_0 yellow (8). In this case, catalytic activity could be detected visually. Soaking E_0 crystals with NADPH changes them from yellow to the brownish color typical for EH_2 . Subsequent soaking with GSSG reestablishes the yellow of E_0 , finishing one cycle of catalysis. Although this simple experiment demonstrates the catalytic competence of the crystalline enzyme clearly, one has to keep in mind that its catalytic rate could be still as slow as the substrate diffusion rate into the crystal. Full equivalence has not been proved.

Even with the advent of Laue-diffraction methods (9), the situation has not changed dramatically. Here, the required synchronization of the reaction throughout the whole crystal introduces new sources of error. Synchronization can be achieved by using light sensitive substrate analogues (10). These are prone to cause active center deformations and/or crystal packing rearrangements that prevent fast measurements.

Apart from enzymes that retain their activity in the crystals, there are also crystalline enzymes that are almost certainly inactive. This applies, for example, for the adenylate kinases, because these enzymes undergo large conformational changes during each catalytic cycle (11) that would break up any crystal.

Deriving Catalytic Data from Structure Analyses

If one is interested in elucidating a particular enzyme reaction, the natural, or preferentially the recombinant enzyme should be isolated and crystallized. This may mean that a number of homologous enzymes of different species have to be scrutinized until a well diffracting crystal form can be produced. The crystallized enzyme species has to be analyzed rather tediously by multiple isomorphous replacement (12), which is usually hampered by the problem of finding suitable heavy atom derivatives that can be located. After establishing the enzyme structure at about 3 Å resolution, further analysis can be done by refinement if native diffraction data up to about 2 Å resolution are available. The refined model can then be used as *Leitstruktur* to solve the structures of all other crystalline enzyme–ligand complexes of this species; either by the difference-Fourier method if the changes are small, or by more elaborate molecular replacement techniques. The analyzed enzyme–ligand complexes should be designed in such a way that the combination of their structures reveals the catalytic mechanism. For recombinant enzymes further support can come from site-directed mutation experiments.

The Catalytic Mechanism of Glutathione Reductase

Sequence and X-ray diffraction analyses were initially done with the human erythrocyte enzyme since this crystallizes well (13). Analyses are now being continued with the homologous enzyme from *E. coli* (14) where a number of most interesting mutants are known (15). The analysis showed that the enzyme is a homodimer with $M_r = 2 \times 52,400$ (human). The subunit can be divided into four domains that are consecutive along the polypeptide chain. The enzyme has two active center regions that are far apart from each other and probably not coupled (16). Each active center is formed by one prosthetic group FAD and residues from five domains, four of which are from one subunit and one from the other. Therefore, a single subunit must be inactive.

In this analysis the *Leitstruktur* was the 1.54 Å resolution structure of the oxidized human enzyme E_0 (17). At this high resolution all non-hydrogen atoms are revealed with a positional accuracy of better than 0.2 Å. In addition, the analysis showed numerous fixed water molecules mostly at the enzyme surface as well as the mobilities (i.e. temperature factors) of all atoms.

The atomic mobilities turned out to be lowest around the flavin moiety of FAD. Here, the temperature factors amounted to only 43% of the overall average. Moreover, there exists a temperature factor gradient along FAD, the values rise from 43% at flavin

to 78% at adenine. A similar gradient was found for the bound co-enzyme NADPH. Here again, the electron transferring nicotinamide moiety is the most rigid part with a temperature factor of 58% of the overall average, while the value of adenine is 99%. The substrate GSSG, however, is not that accurately bound. Its temperature factor amounts to 140% of the average. Thus, the high resolution structure shows that the electron transfer from NADPH to flavin to the redoxactive disulfide of the protein (see below) requires an extremely exact atomic geometry. In contrast, the bridge exchange between the protein and the glutathione disulfides permits a larger range of geometries.

Starting from the *Leitstruktur* of E_0 ; the catalytic cycle was sampled by a number of enzyme-ligands and -modifications (18). The resulting structures were all solved at 2 Å or better resolution with the help of the *Leitstruktur*. It is very advisable to obtain such a resolution in order to establish all atomic positions with an accuracy that is approximately equal to the temperature movements and in order to find the fixed water molecules safely. Tedious analyses to even higher resolution do not seem to be necessary for clarifying the chemical reactions.

The elucidated crystal structures of glutathione reductase are shown in figure 2 under the heading "X-ray". E_0 is the *Leitstruktur* which could be analyzed most easily using the native crystals in a stable form. $EH_2:NADH$ is the abortive complex between the reduced enzyme EH_2 and the reduced co-enzyme NADH. NADH was used for technical reasons: since the nicotinamide-ribose moiety of NADH binds identically to the respective moiety of NADPH, this complex yields genuine information on the catalysis. The stable blocked enzyme $E_1-SCH_2CONH_2$ permitted the convenient measurement of the charge transfer complex between flavin and the thiolate at high resolution. This complex is very similar to the stable intermediate EH_2 , which was only measured at lower resolution. The stable abortive complex between the oxidized enzyme E_0 , the oxidized co-enzyme $NADP^+$ and the oxidized substrate GSSG revealed the exact site of substrate GSSG. The mixed disulfide between protein and glutathione finally showed an intermediate of the reaction, that could be stabilized by raising the pH to 8.6.

Based on these structures, the catalytic cycle of the enzyme has been subdivided into six states *A* through *F* of figure 2. X-ray analyses at medium resolution had already indicated (19) that electron transfer (from NADPH to the protein sulfurs) and disulfide bridge exchange are geometrically well-separated. No ligand at the NADPH-side of flavin causes geometric changes at the disulfide side of flavin, except of course the protein disulfide bridge opening on reduction. This observation has now been confirmed by high resolution structures of $EH_2:NADH$ and of the blocked reduced enzyme $E_1-SCH_2CONH_2$. The former shows no change at the GSSG binding-site while the latter shows no change at the NADPH binding-site. This clear structural separation of the first and the second half of the catalytic cycle helped appreciably in correlating the observed X-ray structures with the catalytic states.

In figure 2 the catalytic state *A* is the oxidized enzyme given by the very well known *Leitstruktur* E_0 . State *B* with NADPH docked on E_0 is an unstable intermediate that proceeds rapidly to state *C* (20). After dissociation of $NADP^+$, state *C* is the stable intermediate EH_2 of the catalytic cycle. The structure of state *B* is derived from those of E_0 and $EH_2:NADH$ assuming that the disulfide bridge opening does not affect the binding of NADH. State *C* (without $NADP^+$) is known from the structures $EH_2:NADH$ and $E_1-SCH_2CONH_2$. State *C* with bound $NADP^+$ is still controversial: The reported full binding of $NADP^+$ in the medium resolution structure of E_1 -

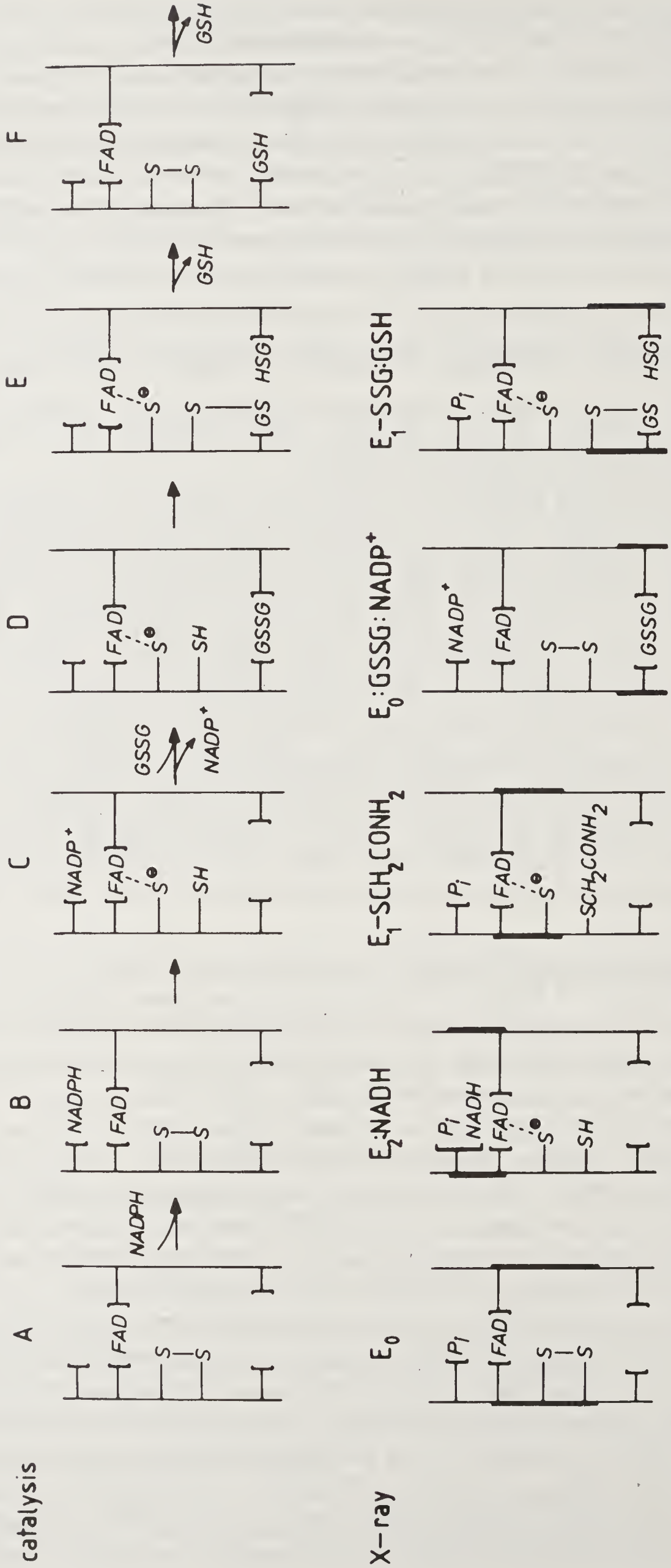


Figure 2. Intermediates of the catalytic cycle of glutathione reductase. Above are the six states A through F, which give a coarse outline of the catalysis. Below are the five X-ray structures, which on combination yield models for all the defined states. The thick vertical bars on the X-ray structures denote the atom arrangement that was taken from the respective analysis. E_0 , E_1 and E_2 denote 0-, 1- and 2-electron reduced enzyme. (Courtesy: P. A. Karplus & G. E. Schulz (1989) *J. Mol. Biol.* **210**, 163–180.)

$\text{SCH}_2\text{CONH}_2\text{:NADP}^+$ (19) was probably an artefact caused by the presence of an appreciable amount of NADPH. Presumably, only the 2', 5'-ADP moiety of NADP^+ binds strongly as observed in the structure $\text{E}_0\text{:GSSG:NADP}^+$ (figure 2). The structure of state *D* can be composed of the bound GSSG taken from $\text{E}_0\text{:GSSG:NADP}^+$ and the reduced enzyme structure EH_2 from $\text{EH}_2\text{:NADH}$. The structure of the mixed disulfide $\text{E}_1\text{-SSG:GSH}$ corresponds directly to the intermediate state *E* of figure 2. It is only stable at high pH values, here 8.6. The structures of $\text{E}_1\text{-SSG:GSH}$ and $\text{E}_0\text{:GSSG:NADP}^+$ can be used to construct state *F*. The catalytic cycle is finished when the second GSH of state *F* has left the enzyme.

After the catalytic states have been correlated with the known structures, the structural changes during catalysis can be specified. In figure 3a the oxidized flavin of state *A* is shown together with the docked reduced nicotinamide of state *B*. Nicotinamide superimposes the central flavin ring so that its hydride donating C4-atom is adjacent to the hydride accepting N5-atom of flavin. The hydride transfer proceeds via an angle $\text{C4-H}\cdots\text{N5}$ of about 135° . Presumably, this angle has approached its optimum during evolution. One can therefore conclude that such an angular hydride transfer is better than a linear one, which solves a chemical query (21). The side view in figure 3a shows that the bound nicotinamide is not quite parallel to flavin and also not planar. Hydrogen-bonds force the carboxamide out of the plane. It should be noted that the enforced carboxamide rotation changes the redox-potential of NADPH.

Bound NADPH transfers the two electrons rapidly to the flavin and then to the

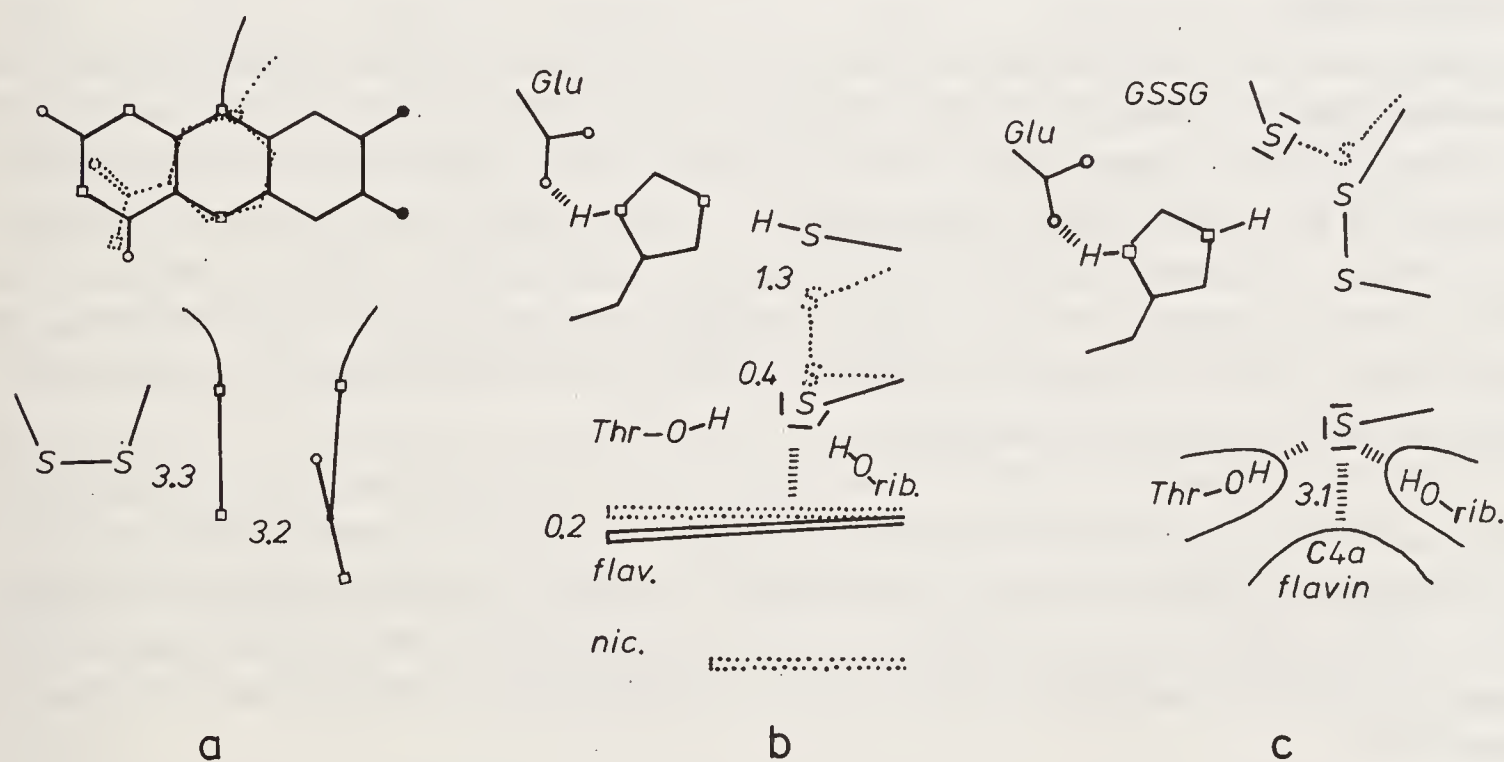


Figure 3. Structural details of the catalytic cycle of glutathione reductase. The atom code is N (\square), O (\circ) and methyl (\bullet). Only some hydrogens are given. All distances are given in Ångströms. (a) Docking of the nicotinamide of NADPH on flavin viewed perpendicularly to the flavin (top, nicotinamide is dotted line in front) and along the flavin ring (bottom, nicotinamide is at the right side, Cys58:Cys63 is at the left side). (b) View from the top of (a) showing nicotinamide, flavin, Cys58:Cys63, Glu472:His467', Thr339, ribityl-O2' of FAD before (dotted lines) and after (solid lines) protein disulfide reduction. (c) Disulfide bridge exchange. Depicted is the mixed disulfide (solid lines) and the sulfur and hydrogen positions of the preceding state consisting of the reduced enzyme and oxidized glutathione.

redox-active protein disulfide Cys58:Cys63 (20). Figure 3b illustrates the electron transfer at the other side of the flavin. It shows the differences between oxidized and reduced enzyme at the disulfide and the catalytic team Glu472':His467' of the other subunit. On reduction, the sulfurs are displaced from a covalent distance of 2.1 Å to a van der Waals distance of 3.8 Å. Spectroscopic data (22) had demonstrated that on reduction Cys63-S γ becomes a thiolate and forms a charge transfer complex with flavin. On complex formation the Cys63-S γ ...flavin distance decreases by 0.2 Å to a final 3.1 Å after flavin has backed up by 0.2 Å. The thiolate is further stabilized by hydrogen bonds from Thr339-O γ and by the 2'-hydroxyl of the ribityl moiety of FAD. Both hydroxyls move by 0.4 Å toward the thiolate to O-H...S hydrogen bond distances of 3.4 Å, which is rather short for such hydrogen-bonds (23). On reduction, the distal sulfur Cys58-S γ is most probably protonated via His467'. Probably, His467' is deprotonated in the oxidized enzyme (24). Cys58-S γ moves by 1.3 Å and forms the catalytic triad Glu472':His467':Cys58 with a good N...H-S distance of 3.4 Å rendering the thiol highly nucleophilic, reminiscent of the active center Ser of the serine-proteases. The enzyme is now poised for disulfide bridge exchange.

This occurs at state *D* where GSSG binds to EH₂ (figure 2). The corresponding structure of figure 3c has been composed from the abortive oxidized complex E₀:GSSG and the reduced enzyme EH₂. A superposition of these structures does not lead to a collision. The sulfurs of Cys58 and glutathione-I are at a short van der Waals distance of 3.1 Å. Cys58-S γ attacks the substrate disulfide and forms a mixed disulfide which is defined here as state *E* (figure 2). The proton at Cys58-S γ is picked up by His467'-N ϵ . The dissected thiolate of glutathione-II is not likely to take the proton from His467'-N ϵ because the geometry is unfavorable. Thiolate is a good leaving group by itself.

In the next step, the proton at His467'-N ϵ polarizes the sulfur of glutathione-I of the mixed disulfide bridge, thus initiating the attack of the stabilized thiolate of Cys63 onto the disulfide. The protein disulfide Cys58:Cys63 is reinstalled leading to state *F* (figure 2). Now His467'-N ϵ can donate its proton in the most favorable geometry to the sulfur of glutathione-I, which dissociates finishing the cycle of reactions. The details of the splitting of the mixed disulfide were essentially derived from the observation of this structure at high pH where His467'-N ϵ is deprotonated so that the attack of the thiolate is inhibited.

The presented catalytic cycle involves merely small atomic movements, which could only be clarified with high resolution structures. The analysis was greatly facilitated by the fact that all events at one side of the flavin are structurally independent from those at the other side of the flavin. In this way, the analysis of the first and the second half of the catalytic cycle can be separated, improving our conclusions appreciably.

The Nucleotide Kinases

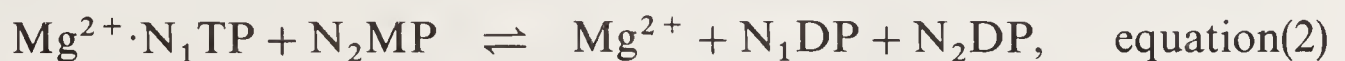
Enzymes that transfer phosphoryl groups are called kinases. Phosphoryls are usually transferred between ester and anhydride bonds. Since the free enthalpy of the anhydride bond is appreciably higher than that of the ester bond, transfer from ester to anhydride is rare, while the other direction is frequent. A particularly labile situation is encountered for a transfer from anhydride to anhydride, because the receiving

substrate can be replaced by the hydroxyl of a water molecule, this being very favorable energetically. This type of transfer is catalyzed by the nucleotide kinases, which therefore have to exclude water from the active center very efficiently.

A similar exclusion of water is also accomplished in glutathione reductase by placing the electron pathway at the molecular center and having the substrates operate from opposite surfaces through the center (figure 2). However, whereas electrons can penetrate a well-packed core, a similar construction for a phosphoryl group would need a wide channel as is, for example, known from the porins (25). But porins exist only in membranes, which facilitate structure- and therefore channel formation. A membrane would require phosphoryl transfer from one compartment to another. Since kinases usually operate within a single compartment, a structural support by a membrane is not possible. Therefore kinases have to be produced as structurally self-supporting water-soluble proteins.

To our knowledge, kinases choose to exclude water by large polypeptide rearrangements during each catalytic cycle, shielding the active center everytime the substrates are bound (11). Most kinases contain more than 350 residues and have therefore enough protection material. The case of hexokinase is well-known (26). The nucleotide kinases are exceptionally small consisting of merely around 200 residues. As a consequence, their relative mass displacements on substrate binding should be even greater than for the large kinases.

The structurally known members of the nucleotide kinase family are shown in figure 4. The catalyzed reaction is



where N_1 and N_2 denote nucleosides, for example, adenosines for the adenylate kinases. All members have closely similar glycine-rich loops indicative of NTP binding (see below). They have also basically the same chain fold. Structural differences occur between the small and large variants. The latter contain an additional domain of 38 residues for shielding the active center. This domain is highly conserved in evolution. For the small variants a 9-residue segment seems to suffice for this purpose. Another remarkable difference was found in guanylate kinase, where the GMP binding domain is formed by a β -pleated sheet instead of α -helices. This may correlate with the fact that guanine is much more water-soluble than adenine and therefore needs stronger interactions for tight binding.

All members of the family are connected by significant sequence homology as indicated in figure 4. Only guanylate kinase has no sequence connection except for the NTP-binding fingerprint sequence. However, the similarity of the chain fold of the main body of the enzyme clearly assigns it to the small variants of the nucleotide kinases.

The structurally best analyzed enzymes are the adenylate kinases, the known structures of which are sketched in figure 5. The first solved structure was that of the cytosolic AK1pig which belongs to the small variants. In its crystalline state AK1pig had no substrates bound to it. The enzyme crystallized in two pH-dependent, interconvertible crystal forms-A (27) and-B (28), both of which were analyzed by X-ray diffraction showing smallish rearrangements. In form-A (high pH) a sulfate binds tightly in a giant anion hole (29) formed by the glycine-rich sequence that has now become the fingerprint sequence for NTP-binding. This sulfate occupies the β -phosphate site of NTP.

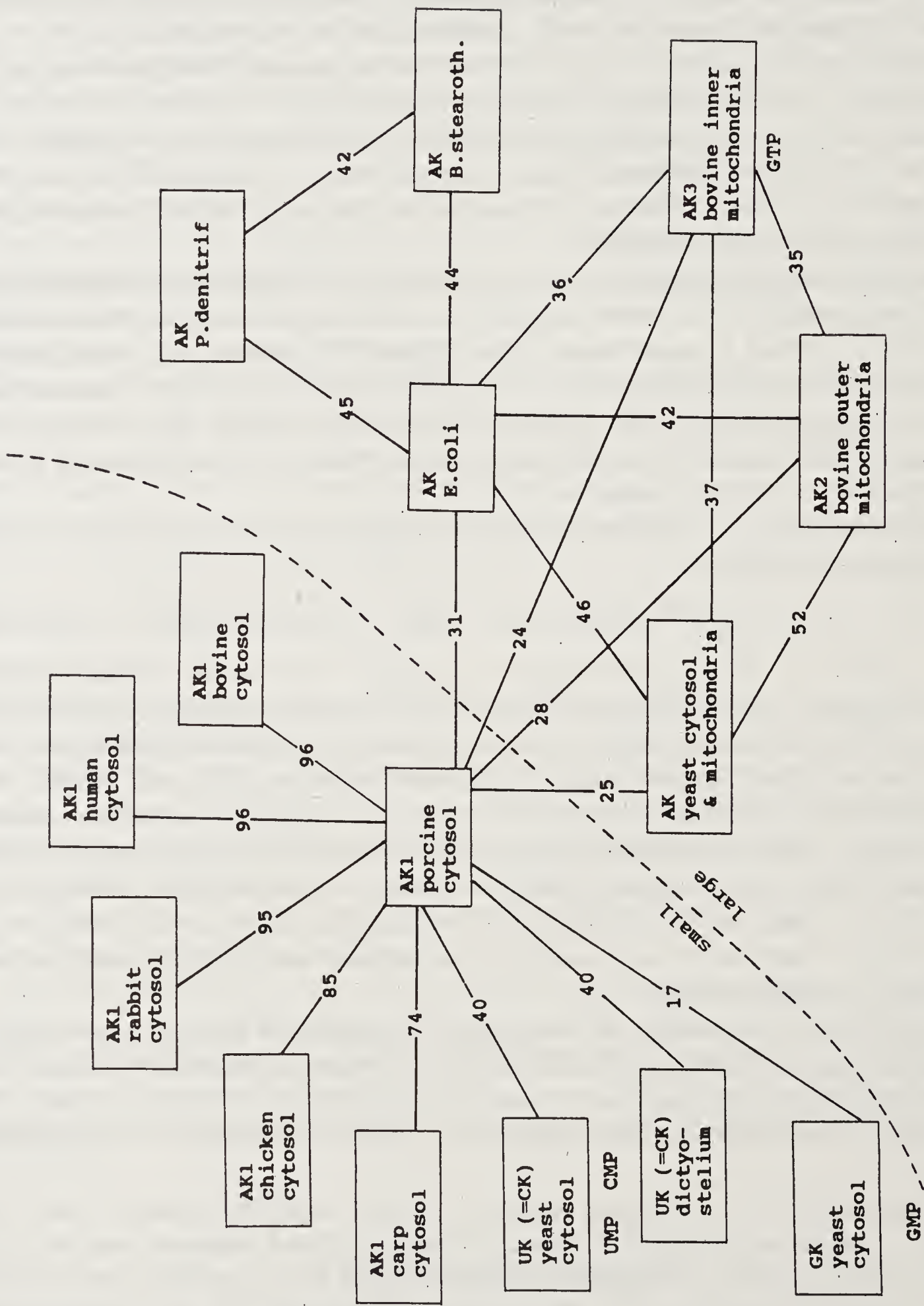


Figure 4. The structurally known nucleotide kinases. AK, UK, CK and GK stand for adenylate-, uridylate-, cytidylate- and guanylate kinase. The small variants are around 21,000 Da whereas the large variants range at 25,000 Da. The percentages of identical amino acids are given. Substrates different from ATP and AMP are indicated.

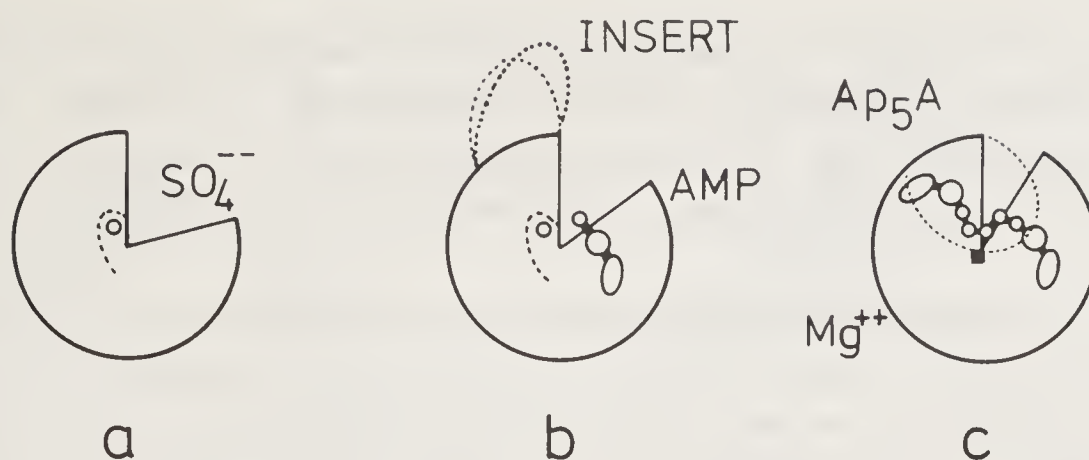


Figure 5. Sketches of the status of the active center cleft in different adenylate kinase crystals (a) AK1 in crystal form-A with bound sulfate at the position of the β -phosphate of ATP (27). The glycine-rich loop forming a giant anion hole (29) is indicated by a dashed line. The same conformation is observed with AK1 carp in a different crystal form (39). (b) AK3 from the bovine mitochondrial matrix co-crystallized with AMP (31, 34). The cleft has decreased, AMP occupies site “adenosine-B”, the sulfate is bound as usual. The location of the domain INSERT is given. There are two AK3 molecules in the asymmetric unit the INSERT domains of which are at slightly different positions. Guanylate kinase binds GMP and sulfate at equivalent positions (35). (c) The structures AKyst:Ap₅A (32) and AKeco:Ap₅A (33). Sites “adenosine-A and -B” are at the left- and right-hand sides of Ap₅A. The Mg²⁺ position has been identified. The domain INSERT (dotted line) covers the active center.

The AK1pig crystals were used for soaking experiments in order to detect the substrate binding sites. These soaks were done with both crystal forms (30). Often they caused crystal breakage or an interconversion between the crystal forms. The experiments yielded the binding site “adenosine-B” as sketched in figure 5. In addition, a second spurious position (30) arose from an undetected crystal-form inter conversion (see ref. 31). Obviously, the enzyme strongly resisted binding any substrate after it was fixed within the crystalline array. Safe information on the binding sites was obtained only much later by co-crystallizations.

Co-crystals were obtained between the adenylate kinase from yeast and the inhibitor P¹, P⁵-bis(adenosine-5'-)pentaphosphate, AKyst:Ap₅A (32). The inhibitor Ap₅A represents ATP and AMP connected by a fifth phosphate and thus mimicks both substrates. Since the difference between ATP and AMP at van der Waals distance (assumed before and after the phosphoryl transfer) and Ap₅A is only 1 Å, the bound inhibitor Ap₅A should be very similar to the Michaelis complex. Indeed, the structure of AKyst:Ap₅A revealed the two binding sites “adenosine-A and -B”, which were later confirmed by the closely resembling co-crystal structure AKeco:Asp₅A of the *E. coli* adenylate kinase (33). Because of the symmetry of Ap₅A, however, these structures failed to yield unambiguous information on which of the two sites accommodates the nucleotide monophosphate and which the triphosphate.

In another structure analysis involving the adenylate kinase from the mitochondrial matrix AK3, co-crystals could be analyzed that contained only one of the substrates, AMP (31, 34). These AK3:AMP crystals showed that AMP binds exclusively at site “adenosine-B”, indicating strongly that this is the monophosphate site. Accordingly, site “adenosine-A” should accommodate the triphosphate. The fewer contacts and the weaker side-chain conservation in evolution at adenosine-A corresponds well with the observed weaker specificity of the adenylate kinases to the nucleotide base of N₁TP as compared to N₂MP of equation (2). The assignment was confirmed by the structure of

GK:GMP (35), i.e. of guanylate kinase co-crystallized with its substrate GMP, which bound at a position equivalent to site “adenosine-B”. A further, more direct confirmation came from the c-H-ras-p21-protein (36, 37), which is structurally homologous to the nucleotide kinases and binds GTP at a position equivalent to “adenosine-A”.

Although N_1 TP forms only weak contacts at its nucleotide base, it is still strongly bound via its α - and β -phosphate as well as by its ribose that extends across the glycine-rich loop. This loop is also present in the mechanical force generating S1-fragment of myosin and in the ATP-generating F_1 -ATPase of mitochondria and chloroplasts and several other nucleotide-using enzymes (29), all of which contain no other sequence-related chain segments. The frequent occurrence of this particular glycine-rich sequence-finger-print and its extremely strong conservation in evolution seems to report a primordial encounter between a nucleotide and a polypeptide, in which a gaint anion hole is formed by the peptide amides. The importance of this binding motif would even increase if the frequent occurrence were not caused by strong conservation in a divergent evolution but by a convergent evolution to a preferable arrangement.

The structure sketches of figure 5 show a large cleft for the initially solved structure of AK1pig (27, 38), which was then intuitively associated with the active center. It contains many positively charged side chains and could thus well accommodate the four phosphoryl groups of the substrates ATP and AMP. Furthermore, sequence comparison demonstrated that the residues lining this cleft are well conserved, corroborating the interpretation. In AK3:AMP the cleft is remarkably smaller (31,34). Almost complete closure of the cleft is observed in AKyst:Ap₅A and AKeco:Ap₅A (32, 33), i.e. when both substrates in the form of the inhibitor Ap₅A are bound. In both steps of this closing procedure, the main body of the enzyme consisting of the central parallel β -pleated sheet and the surrounding α -helices remains intact.

The 30-residue-segment forming the N_2 MP-site undergoes a movement of up to 8 Å on AMP-binding. Moreover, there occur movements of the 38-residue comprising domain INSERT, which is characteristic of the large variants of the nucleotide kinase family (figure 4). In AK3:AMP, this domain idles at the cleft's mouth (figure 5b), but it comes down and shields the phosphates when both substrates are bound as in AKyst:Ap₅A and AKeco:Ap₅A. It is remarkable that the movement of this domain is a pure rotation by about 90° causing main chain displacements of up to 32 Å. This demonstrates that the domain is a solid body i.e. a solid lid that closes down after the substrates are bound and catalysis can proceed. These movements occur during each catalytic cycle, i.e. every 2 ms, if the enzyme works close to its maximum speed.

The catalytic movements are more accurately illustrated in figure 6. Here, the high resolution structures of substrate-free AK1pig, AK3:AMP and AKeco:Ap₅A have been superimposed with respect to the main bodies of the molecules, and then separated by a translation so that the rotation of the lid and the shearing-like movement of the N_2 MP-binding domain can be visualized. When discussing these movements one should take notice of the fact that the active center is disassembled in the substrate-free structure of AK1pig. Only on cleft closure, there emerge aspartates that fix the guanidinium groups of the arginines in the active center, which in turn fix the phosphates. Thus, the enzyme not only undergoes large movements during a catalytic cycle but also assembles and destroys the arrangement of the catalyzing residue team. It takes really great care to avoid ATP hydrolysis.

Because of the large movements one cannot expect that any nucleotide kinase is

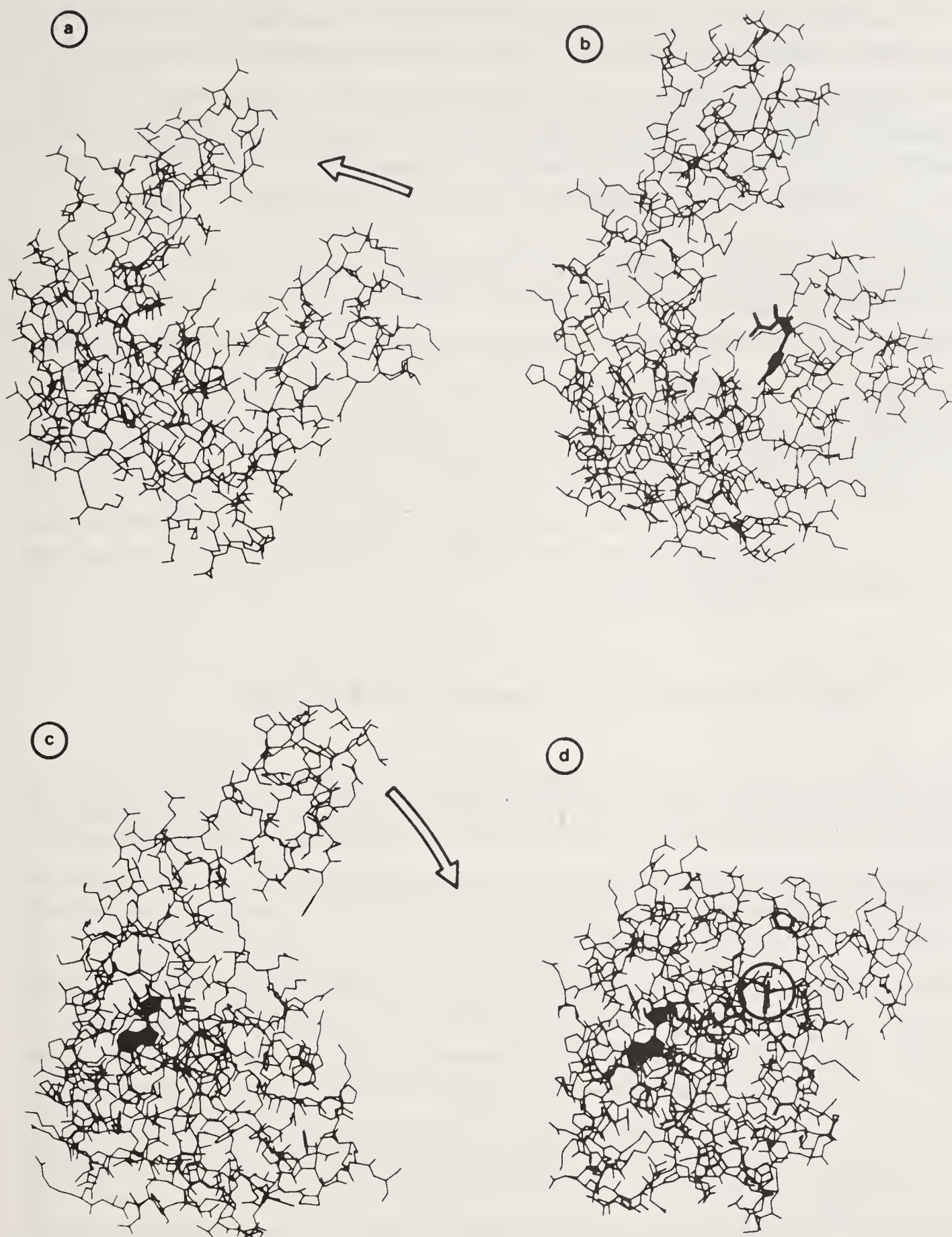


Figure 6. Domain movements correlated with substrate binding in adenylate kinases. The models are given with all non-hydrogen atoms, the substrates are marked. The movements are indicated by arrows. (a) AK1 without substrate (27). (b) AK3 with bound substrate AMP (31, 34). (c) Same as (b) after a rotation of 90° around a vertical axis. (d) AKeco with bound two-substrate mimicking inhibitor Ap₅A (33). The models of (a) and (b) as well as (c) and (d) have the same orientations as defined by the main bodies containing the central parallel β -sheets of the enzymes. (Courtesy: G.E. Schulz, C.W. Müller & K.Diederichs (1990) *J. Mol. Biol.* **213**, 627–630.)

active in a crystal. This implies that the catalytic mechanism can only be elucidated by analyzing a number of different co-crystals. Thus crystal packing contacts freezing out one particular enzyme conformation may become important. One should therefore make sure that the observed enzyme conformations represent intrinsically stable, major conformations in solution. This should be the case if one finds the same conformation in different crystal packings as in the early structure analysis of α -chymotrypsin (6).

With respect to the structures of figure 6, we find the open-cleft conformation of the substrate-free enzyme in AK1pig and in AK1carp in different packing arrangements. In the AK3:AMP co-crystals there are two enzyme molecules in the asymmetric unit that are identical except for a small rotation of the lid-forming domain INSERT as indicated in figure 5b. As mentioned above, two molecules in one asymmetric unit usually have packing contacts as dissimilar as in different crystal forms. The closed-cleft form is found with virtually identical chain folds in three packing environments, two molecules of AKeco:Ap₅A in one asymmetric unit and AKyst:Ap₅A in a different crystal form. The recurrence of these conformations in different packings indicates strongly that they are indeed intrinsically stable.

Two Extremes in the Spectrum of Enzymes

The rather different behavior of the large dimeric glutathione reductases and the small monomeric nucleotide kinases is likely to represent two extremes of the enzyme world. In glutathione reductase there occur only small movements along the path of transferred electrons while there are massive movements in the nucleotide kinases to protect the phosphoryl transfer, in particular the transfer from one labile phosphoryl anhydride bond to another that occurs here.

The different levels of intrinsic movements of these enzymes seem to reflect different levels of protein evolution. Most probably, the kinases belong to the early enzymes as they participate in very basic metabolic pathways like glycolysis. With their large mechanical movements they appear to us as monsters from primitive steam-engine times. At the contrary, glutathione reductase is a younger enzyme that occurred on the emergence of oxygen, which was late in the evolution of molecules. With its elegantly small movements on electron transfer it seems to reflect much more modern, so-to-say silicon-chip times in the evolution of enzymes. As was shown above, it is much easier to analyze the working principle of the "modern" electronic enzyme than of the "primitive" monster, because the best and more or less only available method, X-ray diffraction analysis, is essentially restricted to stable static molecules.

References

1. Howard, A. J., Gilliland, G. L., Finzel, B. C. & Poulos, T. L. (1987) *J. Appl. Crystallogr.* **20**, 383–387.
2. Renetseder, R., Dijkstra, B. W., Kalk, K. H., Verpoorte, J. & Drenth, J. (1986) *Acta Crystallogr.* **B42**, 602–605.
3. Miyahara, J., Takahashi, K., Amemiya, Y., Kamiya, N. & Satow, Y. (1986) *Nucl. Instrum. Meth. Phys. Res.* **A246**, 572–578.
4. Wang, B. C. (1985) *Methods Enzymol.* **115**, 90–112.

5. Brünger, A. T., Kuriyan, J. & Karplus, M. (1987) *Science* **235**, 458–460.
6. Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. (1967) *Nature (London)* **214**, 652–656.
7. Kline, A. D., Braun, W. & Wüthrich, K. (1986) *J. Mol. Biol.* **189**, 377–382.
8. Krohne-Ehrich, G., Schirmer, R. H. & Untucht-Grau, R. (1977) *Eur. J. Biochem.* **80**, 65–71.
9. Hajdu, J., Machin, P. A., Campbell, J. W., Greenhough, T. J., Clifton, I. J., Zurek, S., Gover, S., Johnson, L. N. & Elder, M. (1987) *Nature (London)* **329**, 178–181.
10. Kaplan, J. H., Forbusch-III, B. & Hoffman, J. F. (1978) *Biochemistry* **17**, 1919–1935.
11. Schulz, G. E., Müller, C. W. & Diederichs, K. (1990) *J. Mol. Biol.* **213**, 627–630.
12. Blundell, T. L. & Johnson, L. N. (1976) *Protein Crystallography*, (Academic Press, New York).
13. Schulz, G. E., Schirmer, R. H., Sachsenheimer, W. & Pai, E. F. (1978) *Nature (London)* **273**, 120–124.
14. Ermler, U. & Schulz, G. E. (1991) *Proteins: Struct. Funct. Genet.* **9**, 174–179.
15. Scrutton, N. S., Berry, A. & Perham, R. N. (1990) *Nature (London)* **343**, 38–43.
16. Scrutton, N. S., Berry, A., Deonarain, M. P. & Perham, R. N. (1990) *Proc. R. Soc. London* **B242**, 217–224.
17. Karplus, P. A. & Schulz, G. E. (1987) *J. Mol. Biol.* **195**, 701–729.
18. Karplus, P. A. & Schulz, G. E. (1989) *J. Mol. Biol.* **210**, 163–180.
19. Pai, E. F. & Schulz, G. E. (1983) *J. Biol. Chem.* **258**, 1752–1757.
20. Huber, P. W. & Brandt, K. G. (1980) *Biochemistry* **19**, 4568–4575.
21. Sustmann, R., Sicking, W. & Schulz, G. E. (1989) *Angew. Chem. Int. Ed. Engl.* **28**, 1023–1025.
22. Williams, C. H. Jr. (1976) in *The Enzymes*, ed. Boyer, P. D. (Academic Press, New York, London) Vol. 13, pp. 89–173.
23. Gregoret, L. M., Rader, S. D., Fletterick, R. J. & Cohen, F. E. (1991) *Proteins: Struct. Funct. Genet.* **9**, 99–107.
24. Boggaram, V. & Mannervik, B. (1978) *Biochem. Biophys. Res. Commun.* **83**, 558–564.
25. Weiss, M. S., Kreusch, A., Schiltz, E., Nestel, U., Welte, W., Weckesser, J. & Schulz, G. E. (1991) *FEBS Lett.* **280**, 379–382.
26. McDonald, R. C., Steitz, T. A. & Engelman, D. M. (1979) *Biochemistry* **18**, 338–342.
27. Dreusicke, D., Karplus, P. A. & Schulz, G. E. (1988) *J. Mol. Biol.* **199**, 359–371.
28. Dreusicke, D. & Schulz, G. E. (1988) *J. Mol. Biol.* **203**, 1021–1028.
29. Dreusicke, D. & Schulz, G. E. (1986) *FEBS Lett.* **208**, 301–304.
30. Pai, E. F., Sachsenheimer, W., Schirmer, R. H. & Schulz, G. E. (1977) *J. Mol. Biol.* **114**, 37–45.
31. Diederichs, K. & Schulz, G. E. (1990) *Biochemistry* **29**, 8138–8144.
32. Egner, U., Tomasselli, A. G. & Schulz, G. E. (1987) *J. Mol. Biol.* **195**, 649–658.
33. Müller, C. W. & Schulz, G. E. (1988) *J. Mol. Biol.* **202**, 909–912.
34. Diederichs, K. & Schulz, G. E. (1991) *J. Mol. Biol.* **217**, 541–549.
35. Stehle, T. & Schulz, G. E. (1990) *J. Mol. Biol.* **211**, 249–254.
36. DeVos, A. M., Tong, L., Milburn, M. V., Matia, P. M., Jancarik, J., Noguchi, S., Nishimura, S., Miura, K., Ohtsuka, E. & Kim, S.-H. (1988) *Science* **239**, 888–893.
37. Pai, E. F., Kabsch, W., Krengel, U., Holmes, K. C., John, J., Wittinghofer, A. (1989) *Nature (London)* **341**, 209–214.
38. Schulz, G. E., Elzinga, M., Marx, F. & Schirmer, R. H. (1974) *Nature (London)* **250**, 120–123.
39. Reuner, C., Hable, M., Wilmanns, M., Kiefer, E., Schiltz, E. & Schulz, G. E. (1988) *Protein Sequence Data Anal.* **1**, 335–343.

Structural Perspectives in the Function of Carbonic Anhydrase

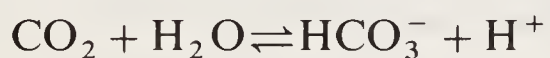
K. K. Kannan

Solid State Physics Division, Bhabha Atomic Research Centre,
Bombay 400 085, India

Introduction

One of the central questions of modern biology concerns the function of enzymes. Though a large number of protein structures are known today, we understand very little about the most fundamental aspect of enzymes, viz. structure–function relationship. The prediction of three-dimensional structure from amino acid sequence is another important problem area which continues to attract the attention of many. Carbonic anhydrase (E.C. 4.2.1.1 carbonate dehydratase) has fascinated me since 1965 when I joined the Protein Crystallographic group of Prof. Strandberg at Uppsala and ventured into macromolecular crystallography (1).

A large volume of literature exists on the biochemical and biophysical studies of carbonic anhydrases and these studies (2–4) have resulted in a number of hypotheses regarding the enzyme catalyzed reaction.



Very little attention has been given to the other “nonphysiological” functions of Carbonic Anhydrase (CA), viz. the ester hydrolysis and hydration of aldehydes.

Carbonic anhydrase is an ancient enzyme. It is found in bacteria, algae, plants, invertebrates and vertebrates. There are 7 distinct forms of isoenzymes known in mammals (5) and one or more of these forms are present in all the different organs of mammals. The recent conference held under the auspices of the New York Academy of Sciences, USA, has in its proceedings a most useful and up-to-date reference book on carbonic anhydrase (6).

Historically the enzyme was first identified in whole blood, purified by Meldrum and Roughton (7) and subsequently shown by Keilin and Mann to be a zinc metalloenzyme (8). Mann and Keilin (9) showed almost simultaneously that sulfanilamide, a drug prevalent at that time, was an inhibitor of carbonic anhydrase which explained its serious side effects due to the fall in the CO_2 combining power of blood of treated patients. Later, other adverse side effects associated with the administration of the drug were also shown to be due to the inhibition of carbonic anhydrase in different organs (2, 6). Thus, the identification of a specific inhibitor of carbonic anhydrase with a drug followed almost immediately after the discovery of the enzyme in the erythrocytes and has attracted attention as a model enzyme for drug design.

Carbonic anhydrases have been shown to be distributed in almost all organs in mammals (4). Most of the reported work on carbonic anhydrase has been carried out in mammals. Many of the isoenzyme forms in mammals have been classified, often, on the basis of their activity and it may well be that a number of these carbonic anhydrases which show high activity may eventually turn out to be different isoenzyme forms of the enzyme, once their amino acid composition and sequence are determined. CA IV, CA V and CA VI are non cytosolic enzymes and have signal sequences which are used for targeting these enzymes to their correct tissue locations (5).

The cDNA sequences of CA I, CA II, CA III, CA IV, CA V, and CA VI isozymes from different species are also known today and consequently their amino acid sequences. The gene structures of CA I, CA II, CA III and CA VII are also known (5) and vary in their size from 9.8 kb for CA VII to 17 kb for the CA II gene. The amino acid sequences have been shown to be coded by seven exons with conserved intervening sequences. CA II gene of mouse differs from that of the human and other CA genes with an upstream displacement of the 4th intron by 14 base pairs. The CA I mouse and human genes show greater variation with an extra intron before the 5' untranslated regions (5).

The availability of cDNA library has also resulted in testing the importance of several amino acid residues in the active sites of CA II, and CA III by site-directed mutagenesis (5, 10). However these studies are yet to reveal the details of the mechanism of carbonic anhydrase function. Other ideas and experiments are therefore required to answer this important question. In this review, the current status of carbonic anhydrase mechanism and protein crystallographic studies as well as the contributions of our group at BARC, Bombay towards these and our thoughts towards the understanding of the structural perspectives of carbonic anhydrase function will be discussed.

Substrates of Carbonic Anhydrase

CO₂ is the only known physiological substrate of carbonic anhydrase isoenzymes from plant to man. The hydration of CO₂ resulting in the formation of HCO₃⁻ is very important in many metabolic pathways, as bicarbonate is a substrate for a number of enzymes in the synthesis of fatty acids, urea, amino acids, carbonic acid, pyrimidine, pyruvate, glucose and glycogen (6). H⁺, the proton secreted in the forward reaction may also be important to maintain the physiological pH. In plants, carbonic anhydrase has been implicated in the photosynthetic fixation of carbon from CO₂ from the atmosphere (3). A cDNA sequence of spinach chloroplast CA with low CO₂ hydration activity reported recently shows no homology to the mammalian enzymes (11) and may be a different group of enzymes unrelated to the known CAs.

Another less understood or indirectly evident function is HCl secretion which has implicated this enzyme in the ulceration of the stomach and duodenum (6). The inhibition of CAI by Cl⁻ anion has also led to the assignment of a possible chloride transport role for this isozyme. The function of a number of CA isozymes such as CAVI in the saliva or CAIII in the red skeletal muscle is not at all understood. The presence of Lys 64 and Arg 67 in the active site, deduced from the structural and sequence homology of CAI, and CAII with CAIII led to the search for phosphatase activity for CAIII, which was subsequently shown for this enzyme (6).

A large number of nonphysiological substrates have been identified for the CA isozymes. In the forward reaction all these carbonic anhydrase catalyzed reactions have H_2O as the second substrate, similar to the CO_2 hydration reaction. Of these the hydrolysis of aromatic and heterocyclic esters has been the most widely studied catalytic reaction of the CAs.

Tables 1a and 1b list the different substrates and inhibitors of carbonic anhydrase. The most important feature of CA catalysis is the different turnover rates for different isoenzymes. CAII is the fastest enzyme known. The turnover numbers for CO_2 hydration are 1×10^6 per sec and 2×10^5 per sec at pH 9 and 25°C , for the CAII and CAI isoenzymes respectively and 10^3 per sec at pH 7.5 for CAIII.

The high turn over rate for the CO_2 hydration for CAII and CAI and the much lower one for CAIII, as well as the low efficiency of the carbonic anhydrases for ester hydrolysis has led to different mechanistic proposals for the different CA catalyzed reactions. Different mechanistic proposals have also been proposed for different CAs. It would be appropriate to postulate a unified mechanism to explain the seemingly different and conflicting catalytic efficiencies of the different CAs. Another intriguing

Table 1a. Reactions catalysed by carbonic anhydrase

1. $\text{CO}_2 + \text{H}_2\text{O}$	$\rightleftharpoons \text{HCO}_3^- + \text{H}^+$
2. <i>p</i> -Nitrophenyl acetate + H_2O	\rightleftharpoons <i>p</i> -Nitrophenol + acetic acid
3. <i>o</i> -Nitrophenyl acetate + H_2O	\rightleftharpoons <i>o</i> -Nitrophenol + acetic acid
4. <i>m</i> -Nitrophenyl acetate + H_2O	\rightleftharpoons <i>m</i> -Nitrophenol + acetic acid
5. Naphthyl acetate + H_2O	\rightleftharpoons Naphthol + acetic acid
6. 2-hydroxy-5-nitro- α -toluene sulfonic acid sultone + H_2O	\rightleftharpoons hydration on the sulfur + ring opening
7. $\text{R}-\text{CHO} + \text{H}_2\text{O}$	$\rightleftharpoons \text{R}-\text{CH}(\text{OH})_2$

Table 1b. Inhibitors of carbonic anhydrase

1. Aromatic or Heterocyclic Sulfonamides
a. <i>p</i> -amino benzene sulfonilamide
b. Acetazolamide
c. Methazolamide
d. Azosulfonamide
2. Anionic Inhibitors
a. F^- , Cl^- , Br^- , I^- , CN^- , SCN^- , N_3^- , HS^- HCO_3^- , MeCO_3^- , $\text{Au}(\text{CN})_2^-$
b. $\text{CH}_3\text{CH}_2\text{CO}_5$, $\text{CH}_3(\text{CH}_2)_2\text{CO}_3$, $\text{CH}_3(\text{CH}_2)_3\text{CO}_3$, $\text{CH}_3(\text{CH}_2)_4\text{CO}_3$, acetate
c. Imidazole, only competitive inhibitor of CAI
3. Metal ions only CAII Cu^{2+} , Hg^{2+}

result of biochemical experiments has been the differing pK_a values for the different enzymes for the group presumed to be associated with catalysis. It is also puzzling that the pK_a changes with buffers used, thus implicating buffers in the catalytic mechanism. The absence of a definite pK_a for CAIII has also raised unanswered questions. Any mechanistic proposal should, no doubt, account for these observations properly.

The structures of the substrates of carbonic anhydrases are also interesting. Except for the sultones, which involve a sulfur–oxygen bond cleavage, all the other substrates involve hydration or dehydration at a carbon.

The simplest of all and the physiological substrate CO_2 is a linear molecule which goes to a planar nonlinear HCO_3^- on hydration. CO_2 has no dipole moment but a quadrupole moment with a residual positive charge on the carbon atom, whereas bicarbonate is negatively charged. The gaseous CO_2 is slightly soluble in water and lipids, and is the metabolic end product of burning, while bicarbonate is very soluble in water.

p-Nitrophenyl acetate ester is an aromatic ring compound with a carbon somewhat similar in bonding to bicarbonate and is hydrolyzed to *p*-nitrophenol and acetic acid while sultone hydrolysis involves only a ring opening (12).

Acetaldehyde hydration is similar in principle in that a water is added to a carbon. In the hydrolysis of CO_2 and aldehydes a carbon–oxygen bond is made, whereas in the hydrolysis of the carbonous esters, a carbon–oxygen bond is broken and in the case of the sultones a sulfur–oxygen bond is broken. It is remarkable that the sultone hydrolysis proceeds orders of magnitude faster than the other aromatic ester hydrolysis. It is also a fact that CAIII is not only a relatively inefficient enzyme of CO_2 hydration but also a poor catalyst of ester hydrolysis. All these observations have to be explained by any viable mechanistic proposal.

Structure of Carbonic Anhydrases

All the physico chemical properties including enzyme catalysis ultimately are interlinked with the molecular structure of these enzymes. The key to enzyme function, be it the obvious commonalities or the glaring differences, arise out of the unique nature of the architecture of the protein, especially that of the active site. The amino acid sequence of a large number of carbonic anhydrases from different species are known as well as the cDNA sequence for some mammalian carbonic anhydrases. We also know the detailed three-dimensional structure of human CAI, CAII and bovine CAII and CAIII (13–18). It is indeed a proud moment in Indian macromolecular crystallography, as has been remarked by Prof. Ramaseshan (19) that this science which was initiated in India by Prof. G. N. Ramachandran has resulted in the first high resolution protein structure of buffalo carbonic anhydrase II (17) with totally indigenous efforts.

A detailed comparison of the structures of HCAI and HCII have been published (13, 16, 20, 21). Here, I will describe briefly the relevant active site environment of the carbonic anhydrase structures. Table 2 lists the residues that are immediately relevant in the active site of carbonic anhydrase isozymes for which we know the primary structure. Some of the residues which are not directly in the active site, are nevertheless, important in providing the proper microenvironment in the active site. It is

immediately apparent from table 2 that there are conserved as well as deviant residues in the carbonic anhydrases. It may be reasonable to expect that the conserved amino acids have functional significance in the nature of catalysis whereas the deviant ones play a role in substrate and inhibitor specificity as well as in the differences in the kinetic properties of the different enzymes. It is important that the similarities and differences in the primary structure as expressed in the 3D structure be correlated for the different isoenzymes for a proper understanding of the structure–function relationship.

The structure of the high activity enzyme, human carbonic anhydrase II (HCAII) was historically the first to be solved at 2 Å resolution (15) by multiple isomorphous replacement method (MIR). On the basis of the known homology in amino acid sequence between HCAI and HCAII, Kannan *et al.* (22) then proposed that the 3D structures of all the carbonic anhydrases would be closely similar to each other. The structure of human carbonic anhydrase I (HCAI) which was subsequently solved by Multiple Isomorphous Replacement method by Kannan *et al.* (13), confirmed these predictions (16). Later the structure of the ‘sluggish’ bovine muscle carbonic anhydrase III (BCAIII) (18) and the fast BCAII (17) from the Indian buffalo erythrocyte were solved by molecular replacement method (MR) and were also found to be very similar to HCAI and HCAII.

The carbonic anhydrases are ellipsoidal molecules of approximate dimension $41 \times 42 \times 55$ Å (20). The structure is formed predominantly by 10 central twisted pleated sheet segments and two minor pleated sheet segments. Four segments of the central twisted pleated sheet structure shape the active site together with loop regions of residues 122–135 and 196–204 (21). 12–16% of the amino acids are in α or 3_{10} helical conformation, which are short and irregular and often have water hydrogen bonded to the carbonyl oxygen thus distorting the helical structure. The pleated sheet structure accounts for about 40% of the residues with some turns, open reverse turns and omega turns forming the loop regions of the structure (figure 1). A 12-Å deep and 12-Å wide mouthed funnel shaped cavity forms the active site of the enzyme to which the essential zinc ion is attached by three histidyl ligands. Another interesting feature of the active site cavity is its division into a hydrophilic half and a hydrophobic

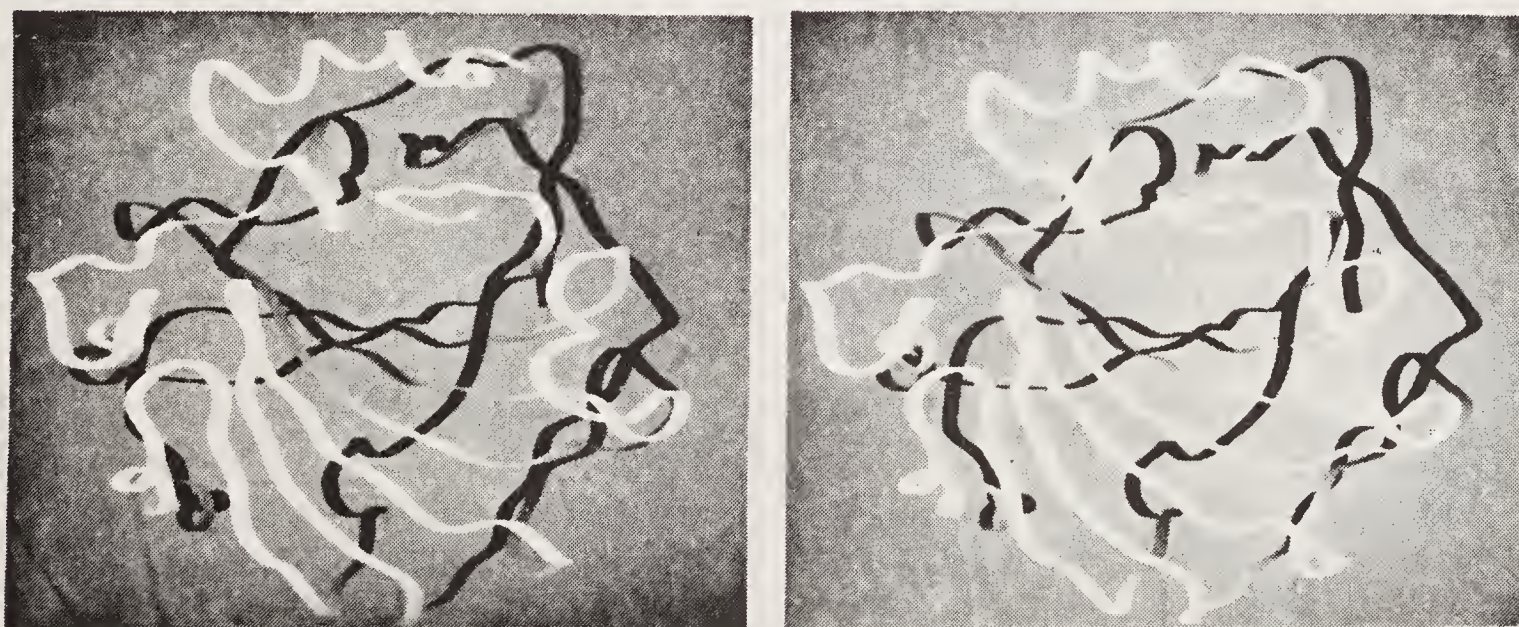


Figure 1. Ribbon drawing of carbonic anhydrase depicting the secondary structure elements. Figure drawn by the program RIBBON developed by Shri Jagdeesh of the Computer Centre, BARC, Bombay.

Table 2. Active site residues of carbonic anhydrases. Zinc ligands are underlined.

Hydrophilic Region															
	7	61	62	64	67	69	92	<u>94</u>	<u>96</u>	106	<u>119</u>	199	200	29	107
	117	194	209												
HCAI	TYR	ASN	VAL	HIS	HIS	ASN	GLN	HIS	HIS	GLU	HIS	THR	HIS	SER	HIS
HCAII	TYR	ASN	ASN	HIS	ASN	GLU	GLN	HIS	HIS	GLU	HIS	THR	HIS	SER	HIS
HCAIII	TYR	ASN	ASN	LYS	ARG	VAL	GLN	HIS	HIS	GLU	HIS	THR	HIS	SER	HIS
BCAI	TYR	ASN	VAL	HIS	HIS	ASN	GLN	HIS	HIS	GLU	HIS	THR	HIS	SER	HIS
BCAII	TYR	ASN	ASN	HIS	ASN	GLU	GLN	HIS	HIS	GLU	HIS	THR	HIS	SER	HIS
BCAIII	TYR	ASN	ASN	LYS	ARG	VAL	GLN	HIS	HIS	GLU	HIS	THR	HIS	SER	HIS
OCAVI	TYR	ASN	ASN	HIS	GLN	SER	GLN	HIS	HIS	GLU	HIS	THR	HIS	SER	HIS
Hydrophobic Region															
	65	91	121	131	135	141	143	145	201	202	204	206	207	211	30
HCAI	SER	PHE	ALA	LEU	ALA	LEU	VAL	GLU	PRO	PRO	TYR	SER	VAL	ILE	PRO
HCAII	ALA	ILE	VAL	PHE	VAL	LEU	VAL	GLU	PRO	PRO	LEU	CYS	VAL	VAL	PRO
HCAIII	THR	ARG	VAL	PHE	LEU	ILE	VAL	GLU	PRO	PRO	GLU	CYS	ILE	LEU	PRO
BCAI	SER	PHE	VAL	PHE	ALA	LEU	LEU	GLU	PRO	PRO	LEU	SER	VAL	ILE	PRO
BCAII	SER	VAL	VAL	PHE	ALA	LEU	VAL	GLU	PRO	PRO	LEU	SER	VAL	VAL	PRO
BCAIII	THR	ARG	VAL	TYR	LEU	ILE	VAL	GLU	PRO	PRO	GLU	CYS	ILE	LEU	PRO
OCAVI	THR	LYS	VAL	TYR	GLN	LEU	VAL	ALA	PRO	PRO	THR	ASN	VAL	VAL	PRO

half, which may be important in enzyme function and in providing the proper microenvironment. A comparison of the α -carbon atoms of HCAI and HCAII (16, 18), of HCAI and BCAII (17) as well as that of HCAII and BCAIII (18) revealed only about 1 Å rms deviation in the respective atomic positions. The different structures had been oriented and positioned properly and the orientation angles and centre of gravity had been refined by Least Squares procedures (23) before carrying out the comparisons (20). The overall shapes of the carbonic anhydrase isozymes from different species are thus very similar. This unique topology has been described as being an unusual “knotty” problem, in that the carboxy terminal region of the structure, which forms an important link to the flanking strand of the central pleated sheet, is tucked inside the protein structure. On pulling this end and the amino end of the protein structure along the direction in which they stretch, a ‘knot’ would be produced in the polypeptide chain (24). This however is an unusual way to visualize protein unfolding, in that one has to imagine nature to be pulling at the ends of the protein structure during unwinding. However, unfolding, denaturation or spontaneous folding during protein maturation is a step by step process and involves the interplay of many amino acid residues and loops of the protein and not by tugging at amino and carboxy ends. One such loop region which will release the carboxy terminus on denaturation, is the 80–86 reverse turn region of CAI, CAII and CAIII, which has incidentally a conserved glycine residue in all carbonic anhydrase isozymes. A deletion of four residues is known in this region for CAVI which has 46 extra amino acids at the carboxyl end. It seems that evolution has taken care of the topological problem in an extraordinary way in carbonic anhydrase CAVI, to preserve the overall folding of the structure, by deleting 4 residues in the 80–86 region such that the extra 46 residues may fold unhindered even while keeping the homologous regions of the carboxy terminus region hydrogen bonded to the flanking region of the protein. Is this an “awareness” at the molecular level or a “chance” trial and error process by which nature has arrived at a shortening of the genomic DNA sequence by 12 base pairs to achieve the removal of 4 amino acids? It would be interesting to devise molecular biology experiments to understand this aspect of protein folding. The tertiary structure of this CAVI would also be interesting and valuable in understanding the evolution of these enzymes. It is pertinent to point out that the placement of the carboxy terminus region in the structure which forms the linking chain segment in the 10-stranded central pleated sheet structure, is essential for the correct folding and preservation of the functionality of the carbonic anhydrase structures (25).

The structure is not only stabilized by the hydrogen bonds of the secondary structure elements but also by amino acid side chain interactions. There are two aromatic clusters in the four known isoenzyme structures with a high degree of homology in the amino acid residues. The exception is that of BCAIII where three phenylalanine residues in one of the aromatic clusters in the interior of the protein are exchanged by Cys, and Leu (18) and two are replaced by Val and Leu in CAVI. However these changes are concentrated in one part of the cluster only thus preserving the character of this aromatic packing to a large extent. Another stabilizing factor is the hydrogen bond network of side chains. Zinc liganded His 119 is hydrogen bonded to Glu 117, which is H-bonded to N 107 and His 107; His 107 to Tyr 194 and Tyr 194 to Ser 29. Ser 29 is hydrogen bonded to N 198 and Trp 209. All of these amino acids are buried inside the protein structure and are also conserved in all the carbonic anhydrases for which sequences are known. The residue His 122 is hydrogen bonded

to Tyr 51 in CAI, CAII and CAIII and can be expected to add to the stability of the structure. Tyr 51 is however replaced by Gly 51 in CAVI as per the published sequence alignment (26), though there is a Tyr in position 52 which would be located on the surface of the protein and unable to participate in this hydrogen bond network. A realignment of the CAVI sequence would probably bring this Tyr 52 in to H bonding position to His 122 thus increasing the homology and preserving the structural stability. Model building studies in the absence of the 3D structure determination may reveal the stabilizing force for the buried hydrophilic His 122. It is interesting that there are two extra residues in this position of the sequence in CAVI, which occur on the surface of the protein and a realignment would not be impossible as glycine, present here, can span the whole of the Ramachandran (ϕ, ψ) space (27).

The active sites of the enzymes are centred around the essential Zn^{2+} , the replacement of which with other divalent transition metals, Hg^{2+} or Cd^{2+} reduces or abolishes the activity. Only Co^{2+} has 50% of the activity of the native enzyme. Zn^{2+} has been shown to be liganded to His 94, His 96 and His 119, all invariant residues, and to a water or OH^- , hydrogen bonded to Thr 199 which is hydrogen bonded to Glu 106. Glu 106 which is buried in the structure is H bonded to a water molecule which is H bonded to Tyr 7. All these residues are known to be invariant in the carbonic anhydrases. Interestingly, even the water molecule is found in all the four carbonic anhydrase structures in the same location as though even this is a sort of a conserved residue. The zinc ligand His 94 is hydrogen bonded to Gln 92 located in the active site, and these are also invariant residues in the CA's.

His 64 in CAI and CAII is replaced by Lys in CAIII. His 67 in CAI is replaced by Asn in CAII, by Arg in CAIII and by Gln in CAVI. Ala 121 and His 200 in CAI are replaced by Val and Thr respectively in all the other CA's. Phe 91 and Leu 131 in CAI are not at all conserved in the other CA's. These changes in the amino acid residues in the active site may reasonably be expected to bring about differences in the catalytic efficiency and strength of inhibitor binding. Two proline residues in the active site form a reverse open turn. Pro 201 and Cis Pro 202 produce the change of direction of the polypeptide chain above the active site and are conserved. A change of this Pro-Pro sequence may have deleterious effect on the structure. Cis Pro 30 which is far removed from the active site, also conserved in the carbonic anhydrases, is located in a turn. However conserved Ser 29 is involved in the H bond network in the interior of the protein molecule described above and Cis Pro 30 can be expected to be important for the stability and integrity of the structure in this region.

Inhibition of Carbonic Anhydrase

Aromatic or heterocyclic sulfonamides with unsubstituted sulfamido groups are specific and powerful inhibitors of the carbonic anhydrases. Table 3 lists some of the sulfonamides that have been complexed with carbonic anhydrase whose structures have been investigated (29–31). Maren has reviewed the chemistry, physiology and inhibition of the sulfonamides in detail and listed the inhibition potency of different sulfonamides (2). Of the different isozymes investigated CAIII is poorly inhibited by the sulfonamides. It is interesting that CAIII is also a poor catalyst of the hydrolysis of *p*-nitrophenyl acetate and other aromatic and heterocyclic esters. The sulfonamides are competitive inhibitors of ester hydrolysis and HCO_3^- dehydration reaction and are generally

Table 3. Sulfonamide inhibition of carbonic anhydrase

Sulfonamide	HCAI		HCAII	
	X-ray resolution	$I_{50} \times 10^7 \text{ M}$	X-ray resolution	$I_{50} \times 10^7 \text{ M}$
Benzene sulfonilamide	2.0 Å	257	—	24
Acetoxy mercury benzene sulfonilamide			2.0 Å	
Acetazolamide	2.0 Å	2.7	3.0 Å	0.02
Methazolamide	2.0 Å	10.0	—	0.16
PL13580	Model building	0.01		0.004

non-competitive to the CO_2 hydration reaction, though shown to be competitive under certain experimental conditions (Lindskog, personal communication). The structures of the HCA I complexes (30, 32) of acetazolamide, methazolamide and benzene sulfonamide as well as the structures of HCAII complex of acetazolamide and acetoxymercury sulfonilamide (18) have shown the sulfonido group to be in the immediate vicinity of the essential Zn^{2+} , replacing the 4th coordinated OH^- or H_2O as well as forming a hydrogen bond with the OG1 atom of Thr 199. It seems very probable that the NH^- group of the sulfonido group binds to Zn^{2+} and H bonds to Thr 199. One of the two oxygen atoms is located at about 3 Å distance to Zn^{2+} at a potential fifth coordination site. This distance is too large for inner sphere coordination to Zn^{2+} and is sufficiently short for an outer sphere coordination. This position corresponds to the binding site of imidazole, the only competitive inhibitor of CO_2 hydration of HCAI (33, 34). Liang and Lipscomb (35) have reported theoretical self consistent field molecular orbital calculations for sulfonamide binding to the carbonic anhydrases, wherein they found a $\text{Be}^{2+}-\text{O}$ distance of 2.56 Å and $\text{Be}^{2+}-\text{N}$ distance of 1.64 Å which are close to the values reported for HCAI sulfonamide complexes on the basis of electron density maps calculated with MIR phases and that of the refined structure of HCAII sulfonamide complexes. There are minimal changes in these values as compared to the refined structures of HCAI–acetazolamide, HCAI–benzene sulfonamide and HCAI methazolamide complexes (unpublished results of S. Chakravarty, Padma Satyamurthy and K. K. Kannan). Liang and Lipscomb also reported that their calculations of pyruamidate ion binding to CAII yields a $\text{Be}^{2+}-\text{O}$ distance of 2.99 Å which they termed as weak binding at the fifth coordination site of the metal ion (35). Kannan *et al.* had termed this a bidentate coordination of the sulfonamides using the same distance criteria (29). This consensus on sulfonamide binding has thus been arrived at, both after refinement of the X-ray coordinates of the complexes and theoretical calculations of metal (Be^{2+} instead of Zn^{2+}) sulfonamide complexes. Thus there is a broad agreement for the proposal put forward first by Kannan *et al.* (34), for a possible fifth coordination site at the essential metal.

Anionic inhibitors behave differently from the sulfonamide inhibitors of carbonic anhydrase. Structures of HCAI–iodide $^-$ (36), HCAI– $\text{Au}(\text{CN})_2^-$ complexes (Vinay Kumar and K. K. Kannan, unpublished results), HCAII–iodide $^-$, and HCAII– $\text{Au}(\text{CN})_2^-$ complexes (31) as well as HCAII– SCN^- (18) complexes are known at high resolution.

Of these, the iodide anion inhibits the enzyme by replacing the 4th coordinated OH^- or water, whereas $\text{Au}(\text{CN})_2^-$ binds in the vicinity of the fifth coordination site while SCN^- binds tightly to Zn^{2+} at the fifth coordination site without either of these replacing the $\text{OH}^-/\text{H}_2\text{O}$ coordinated at the 4th coordination site of the metal. The mode of binding of $\text{Au}(\text{CN})_2^-$ and SCN^- are however different in HCAII, though $\text{Au}(\text{CN})_2^-$ binding is very similar in HCAI and HCAII. Inhibitor binding does not change the hydrogen bonding network of Thr 199–Glu 106–water–Tyr 7. This may have significance in that the sulfonamides have been shown to bind first as neutral species to the enzyme and then get ionized to $-\text{SO}_2\text{NH}^-$ upon binding (37) with a proton transfer from the sulfonamido group akin to the proton transfer observed in the enzyme catalyzed hydration reaction.

The Mechanism of Carbonic Anhydrase

The simple reaction, that of hydration of CO_2 to bicarbonate has occupied a place of preeminence in the mechanistic literature of carbonic anhydrase ever since this enzyme was discovered. This interest is mainly for the following reasons.

1. This is a deceptively simple reaction and the substrate molecule CO_2 and water are very tiny compared to the huge size of the enzyme of 30,000 daltons.
2. The uncatalyzed reaction at pH 7 is orders of magnitude slower than the enzyme catalyzed one (38).
3. The end product of the hydration reaction is bicarbonate which means that a proton transfer has to take place every micro second for the CAII enzymes with a turn over rate of 1×10^6 CO_2 molecules/sec/enzyme molecule and the CAI enzymes coming a close second with 5×10^5 CO_2 molecules/sec/enzyme molecule.
4. The CAI and CAII enzymes representing the faster carbonic anhydrases whose kinetics have been extensively investigated are reported to be pH dependent for V_{max} , the substrate saturation parameter, while the Michaelis constant K_m is independent of the pH. The pH rate profile has been shown to be sigmoidal with an inflection at pH 7 giving a pK_a of 7 for the activity linked group, at least in some buffers which include sulfates. In the absence of such buffers two inflection points were found, one with a pK_a of 6 and another with a pK_a of 9.5.
5. The larger substrates like esters are hydrolysed at a much lower rate than CO_2 hydration.
6. The anionic inhibitors are non/uncompetitive inhibitors of the hydration reaction but competitive to the dehydration reaction and ester hydrolysis.
7. Sulfonamides are competitive inhibitors of the dehydration reaction and ester hydrolysis and under preincubation before CO_2 addition are competitive to the hydration reaction.
8. Most importantly, the apoenzyme is totally inactive, as also most of the enzymes with divalent metal ions substituting Zn^{2+} , except Co^{2+} which has 50% activity and the Cd^{2+} enzyme which has its pH rate profile shifted to higher pH as compared to the Zn^{2+} enzyme.
9. The binding site of CO_2 and bicarbonate in the enzyme during catalysis has not been directly established with certainty by any of the physico-chemical methods.

A number of mechanistic proposals have been made on this basis by different

investigators. There is however a consensus that a zinc bound water going over to an OH^- , is directly responsible for the hydration/hydrolysis reaction. The details of the mechanism, especially the important proton transfer step and the substrate/product association/dissociation steps differ, even though some convergence of thoughts is taking place recently especially around the proposal originally made by Kannan *et al.* (34).

Lindskog and coworkers and others (3, 39–47) proposed that the activity linked group was the OH^- liganded to the metal ion first proposed by Davis (45) as well as Devoe and Kissianowsky (46). The problem that $\text{Zn}^{2+} \text{OH}^-$ was known to have a very high pK_a (47) above 9.5, has been ignored in these proposals. Another deficiency of this mechanistic proposal was its silence on the protein groups involved in the mechanism except for ZnOH^- . No mention was made of the groups responsible for internal proton transfer first reported to be important by Tu and Silverman from ^{18}O exchange studies (48), except as a mere buffer mediated step. However the final proton transfer group was identified as His 64 present in CAII and shown by NMR studies (49) to have a pK_a of 7.0 as required by the pH rate profile of catalysis. This His 64 is not present in CAIII and is present in CAI but is inaccessible in the active site due to the shielding by His 67 and His 200. A number of water molecules were shown (40, 42, 50, 51) to be interspersed between the Zn^{2+} bound water and His 64. This proposal received a very severe set-back when Lindskog and coworkers (10) showed by site-directed mutagenesis that this residue is not involved in the activity of the enzyme. Recent ^{18}O exchange studies of Ala 64 CAII (52), have however shown that His 64 can be involved in the proton transfer function but unimportant for catalysis. Surprisingly imidazole buffer can carry out the function of His 64 in its absence in CAII. This proposal will however require different mechanistic proposals for different carbonic anhydrases, notwithstanding the high primary and tertiary structural homology of the isoenzymes. It is interesting that a known inhibitor of CAI is an activator of Ala 64 CAII. The buffer mediated proton transfer seems to be valid only in imidazole buffer raising the question of the binding site of imidazole *vis-a-vis* other protein groups to effect the internal proton transfer. Lastly, the term “internal proton transfer” will become a misnomer as the proton transfer will be through solvents in the active site and not through any protein groups. Thus the results have raised more questions on the groups involved in the catalytic mechanism than providing answer to any of them conclusively.

A different proposal by Appleton and Sarkar (53) and later by Pesando (54) invoked one of the histidine ligands of the essential Zn^{2+} ion as the activity linked group on the basis of NMR studies which was later shown to be untenable by Bruce Martin (55) and also by Coleman and Lindskog (56).

NMR and EPR studies as well as other spectroscopic studies have been used to pinpoint the substrate binding site with respect to the essential metal ion. Riepe and Wang first proposed from Infrared (IR) studies that CO_2 is bound in a hydrophobic pocket without appreciable strain and that it was not coordinated to the metal ion (57). Bertini and Luchinat (58) as well as Yeagle *et al.* (59) inferred from ^{13}C NMR studies that HCO_3^- binds to the metal ion with the ^{13}C located about 3.5 Å, while CO_2 was located at about 8 Å distance from the metal ion. An important contradiction in this proposal is the negation of microscopic reversibility of the hydration/dehydration reactions. In one step i.e., hydration reaction, the CO_2 , which is located 8 Å away from Zn^{2+} will get converted to HCO_3^- by water mediated catalysis involving more

than one water molecule in H bond network from the metal liganded OH^- ion or the bicarbonate itself, serving as the proton transfer group, whereas the end product of the catalysis i.e. HCO_3^- will be found bound to the metal ion. In the dehydration step the distance relation between the substrate HCO_3^- and H^+ and the end products CO_2 and H_2O will have similar geometries thus keeping the site of binding of the products CO_2 and water far removed from the site of the substrate HCO_3^- necessitating a hop-step-and-jump of substrates and products. Thus even the so-called direct dynamic studies of NMR, IR and EPR have given contradictory information on the localization of the substrates as well as the mechanism of the important proton transfer step.

Kannan *et al.*, for the first time proposed a probable fifth coordination binding site for CO_2 (34), on the basis of X-ray diffraction studies of the inhibitor complex of HCAI with imidazole the only competitive inhibitor of CO_2 hydration reaction as shown by Raja Khalifha (33) in Prof. Edsall's laboratory at Harvard, USA. Kannan *et al.*'s proposal (34) at that time did not involve any histidine residue in the proton transfer step even though the refined structure of HCAI (14) has now shown an extension of the hydrogen bond net work from Glu 106 to His 64 and His 200 through a conserved water molecule, Tyr 7 and other water molecules in the active site. As already mentioned Thr 199 is hydrogen bonded to the fourth liganded water/ OH^- of zinc and to the buried Glu 106. These two amino acid residues are conserved in all the carbonic anhydrases for which sequences are known and are situated right in the activity centre. The inhibitor imidazole, whose structure was interpreted from the electron density maps calculated with MIR phases and with $(2F_{\text{imidazole}} - F_{\text{native}})$ and $(F_{\text{im}} - F_{\text{nat}})$ as coefficients revealed the position of the imidazole molecule near zinc in a different position compared to the NH^- group of sulfonamides and iodide but more resembling the binding of $\text{Au}(\text{CN})_2^-$ binding. The 4th liganded water molecule was also not replaced by this inhibitor. Model building showed that the plane of the imidazole ring was not directly pointing to the zinc ion. The nitrogen of the imidazole ring was placed at about 2.7 Å distance from Zn^{2+} . The $\text{Au}(\text{CN})_2^-$ molecule partially overlaps the imidazole position but does not extend to the zinc ion. The nitrogen of the imidazole was also found to be almost overlapping one of the oxygen positions of the sulfamido group proposed by Kannan *et al.* (29, 30) to be at a fifth coordination site. This oxygen was also measured to be at about 2.8 Å from the difference electron density maps calculated with MIR phases.

The refinement of protein structures were not considered feasible at that time, after initial attempts at Cambridge to refine myoglobin structure were not fruitful. With the advent of restrained/constrained Least Squares methods (60,61) and computer graphics (62) it has now become possible to refine protein structures and get meaningful results. The structures of HCAI native (14) and sulfonamide complexes of HCAI (30) and HCAII and native HCAII (18) have now been refined and these interpretations have been confirmed. Erickson *et al.* (31) found for the HCAII-acetoxy mercury sulfonilamide complex, a Zn^{2+} -oxygen distance of 3.1 Å and interpreted this to be outside the coordination sphere of zinc giving it only a tetrahedral coordination instead of the penta coordination proposed by Kannan *et al.* (29), with one of the oxygens of sulfamido group forming the distant fifth ligand.

Recently Liang and Lipscomb (35) and Vedani *et al.* (51) have on the basis of quantum chemical calculations, proposed a five coordination geometry for sulfonamide binding, very similar to that of Kannan *et al.* (29), with almost identical distance

geometries as that of Erickson and Liljas (18). The SCN^- inhibited HCAII enzyme has also been shown to exhibit five coordination geometry with some rearrangement of the 4th liganded water/ OH^- (18).

Thus, on the basis of the structure of the refined HCAI (14) and HCAII (18) structures and their inhibitor complexes a consensus has emerged as to the possible substrate binding site near the zinc ion, most probably the distant fifth coordination site first proposed by Kannan *et al.* (35).

Figure 2 shows a schematic of the active site of HCAI with solvent molecule positions identified from difference Fouriers calculated with the refined phases, and refined to an R factor of 17.8% at 2 Å resolution. A closely similar water structure has been found in HCAII also (18) up to His 64.

The structure has also shown a possible proton transfer pathway (14, 35). The near neutral pK_a of the activity linked group had for a long time favoured a suitably located histidine residue in the active site as the preferred group for this proton transfer function to buffer. However, as already mentioned, the site directed mutagenesis of His 64, the lone active site candidate in HCAII, showed that the catalytic efficiency of the enzyme is unaffected by the substitution of His 64 by other amino residues (10). Other proposals have involved the bicarbonate hydrogen bonded to water molecules in the active site (59). The only proposal, based on the structural studies of HCAI and HCAII for the proton transfer pathway that did not include a histidine residue was that of Kannan *et al.* (34). The Zn-OH-Thr 199-Glu 106 proton shuttle was put forward on the basis of the structural studies of imidazole HCAI complex and known amino acid homology of other carbonic anhydrases. Early theoretical calculations of Pullmann on this proposal (63) showed that there was no energy

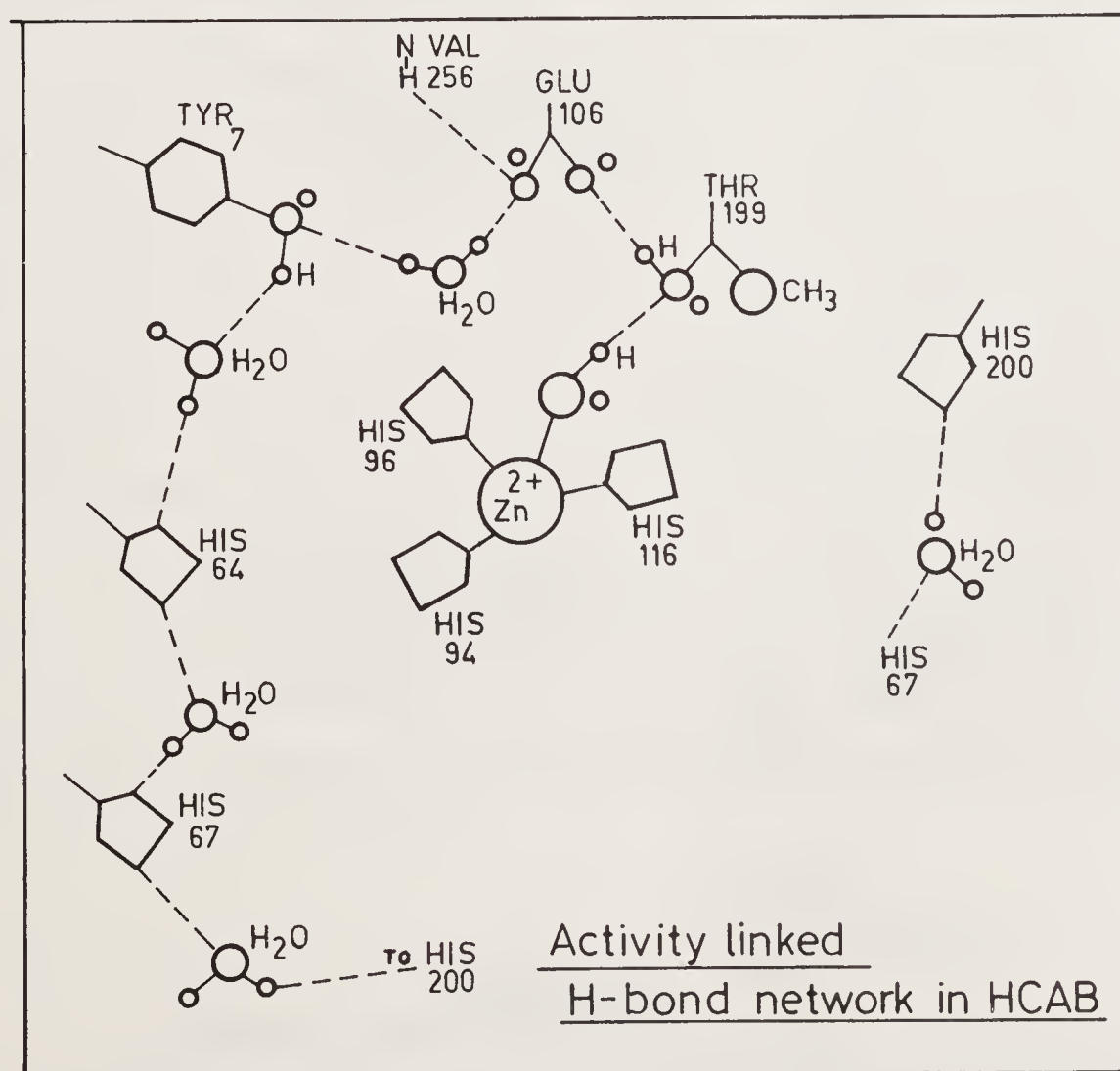


Figure 2. Schematic representation of some of the active site residues and associated solvent structure.

barrier for this proton shuttle to function efficiently. Recently Merz *et al.* (50) also found that a proton gets transferred to Glu 106 spontaneously without any energy barrier through the proton shuttle and as such did not find any theoretical grounds to reject this mechanism. Kannan *et al.* (14) have shown recently that the H bond network can be extended to His 64, His 67 and His 200 in HCAI.

Thus, the proton shuttle of Kannan *et al.*, is the only one which has been discovered from structural studies and is still a tenable pathway of internal proton transfer in carbonic anhydrases. This mechanism is also consistent with ester hydrolysis (figure 3) and aldehyde hydration reactions of CA's.

This mechanism involves the CO_2 binding as a weak ligand to a distant fifth coordination site of the metal ion similar to the imidazole. This site is located in the hydrophobic half of the active site, thus correlating well with the infrared studies (57). This is also in agreement with the ^{13}C NMR studies (40, 59) where the HCO_3^- was shown to bind in the inner coordination sphere and CO_2 near the vicinity of the metal ion. The water which binds at the 4th coordination site transfers a proton

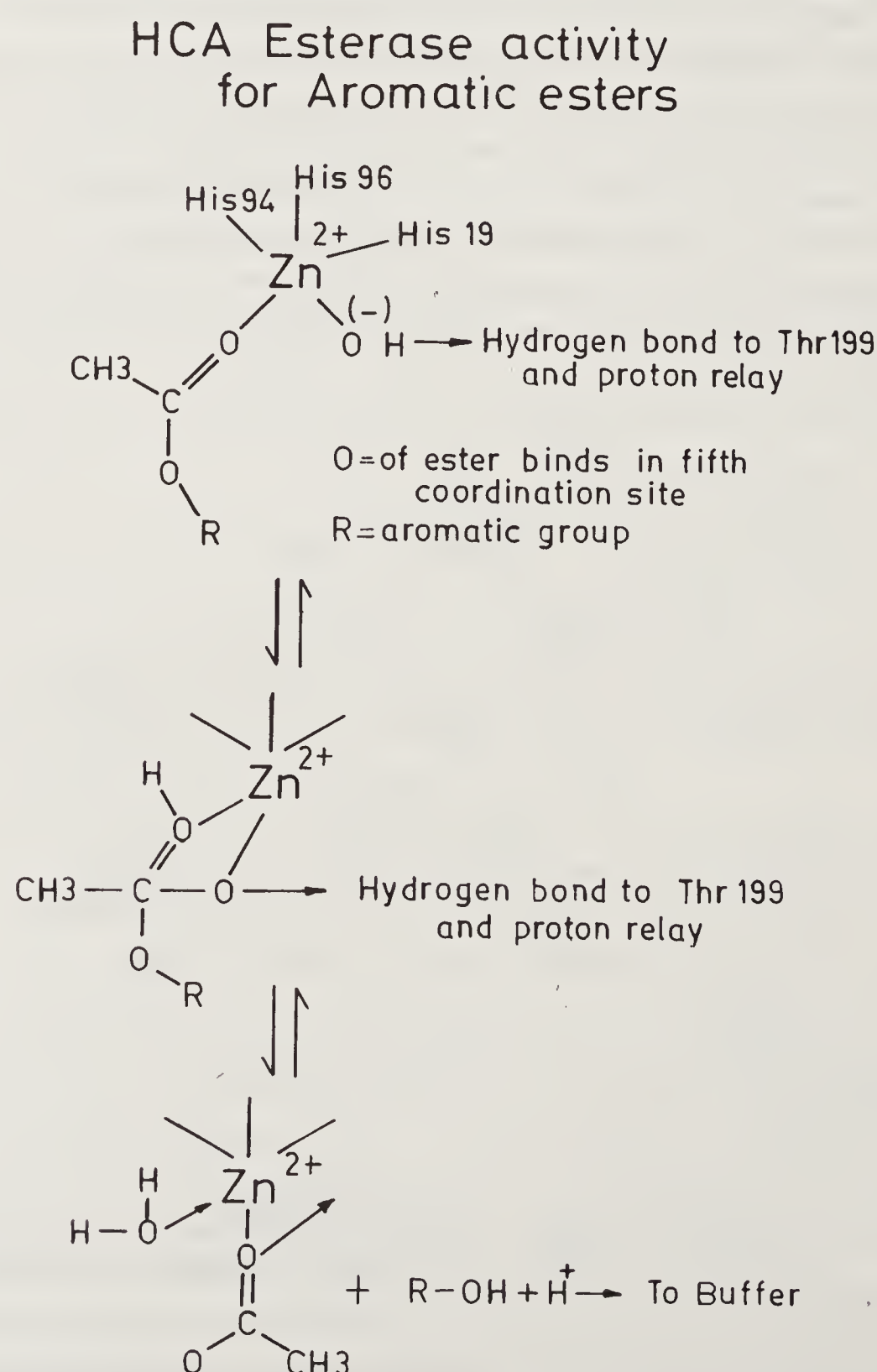


Figure 3. Ester hydrolysis pathway in carbonic anhydrase.

to Glu 106 through the proton shuttle to become OH^- . Quantum chemical calculations by Pullman and Merz *et al.* (63, 50) have shown this step to be spontaneous and without requiring extra energy. The OH^- thus formed is positioned at the right orientation and proximity to the fifth liganded CO_2 to effect a nucleophilic attack on the positively charged carbon atom forming a transition state going to the product bicarbonate which is displaced by a water molecule to start a new cycle after the proton has been transferred to the buffer.

Site directed mutagenesis will verify the importance of Thr 199, Glu 106 and other protein groups including the zinc ligands in the activity of carbonic anhydrases. We have undertaken these studies as well as substitution of other amino acids by genetic engineering experiments, which are expected to throw light on the nature of the active site and structure–function relationship. These studies will also be useful in drug design for this clinically important enzyme. In these experiments we will be using the active site perspectives resulting from the 3D structure as determined by X-ray diffraction analysis to design proper protein engineering experiments. We are also initiating studies on a large number of carbonic anhydrase inhibitor complexes and metal ion replacements to further our understanding of enzyme catalysis. We are also planning structural studies on other carbonic anhydrase isozymes from different species to understand the evolutionary aspects of this enzyme.

Acknowledgement

The summer of 1960 was a memorable event in my life as it happened to be my first visit to a research laboratory in India. My B.Sc Physics teacher, Shri Anantha Krishnan of St. Josephs College, Trichy, Tamilnadu, was engaged in research in Prof. G. N. Ramachandran's laboratory. I eagerly accompanied my father Shri K. S. Kothandapani Iyengar, a teacher at the same college, when he went to meet his old friend. There was pride in Shri Anantha Krishnan's face that he was doing research in Prof. G. N. R.'s laboratory. That was how I came to know of the name of one of our greatest scientists G. N. R., as he was popularly and affectionately known and continues to be known even today. I have met Prof. Ramachandran many times during my research career, often as a visitor and have received encouragement from him. My eagerness to search for and take up protein crystallographic research, has in no small way, been influenced by his personality. It is a great pleasure to dedicate this paper to Prof. G. N. Ramachandran.

It is a pleasure to acknowledge the many useful and stimulating discussions which I had with Dr. R. Chidambaram. I have had many fruitful discussions with my colleagues Dr. M. Ramanadham, Dr. V. S. Yadava, Dr. M. V. Hosur, Shri. Vinaykumar, Shri S. Chakravarty and Smt Padma Satyamurthy during the course of the work on carbonic anhydrase.

References

1. Fridborg, K., Kannan, K. K., Liljas, A., Lundin, J., Strandberg, B., Strandberg, R., Tilander, B. & Wiren, G. (1967) *J. Mol. Biol.* **25**, 505–516.
2. Maren, T. H. (1967) *Phys. Rev.* **47**, 595–781.

3. Lindskog, S., Henderson, L. E., Kannan, K. K., Liljas, A., Nyman, P. O. & Strandberg, B. (1971) in *The enzymes*, 3rd edition, ed. Boyer, P. D. (Academic Press, New York), **5**, 587–665.
4. Bertini, I., Luchinat, C. & Scozzafava, A. (1982) *Struct. Bonding* **48**, 45–92.
5. Edwards, Y. (1990) *Biochem. Soc. Trans.* **18**, 171–175.
6. Tashian, R. E. & Hewett-Emett (eds). *Annals. N. Y. Acad. Sci.* (1984), **429**, The whole volume p. 1–840 reviews the developments upto 1984.
7. Meldrum, N. U. & Roughton, F. J. W., *J. Physiol (London)* (1933) **80**, 113–142.
8. Keilin, D. & Mann, T. (1940) *Biochem. J.* **34**, 1163–1176.
9. Mann, T. & Keilin D. (1940) *Nature (London)* **146**, 164–165.
10. Forsman, C., Behravan, G., Jonson, B., Liang, Z., Lindskog, S., Ren, X., Sanstrom, J. & Wallgren, K. (1988) *FEBS Lett.* **229**, 360–362.
11. Burnell, J. N., Gibbs, M. J. & Mason, J. G. (1990) *Plant Physiol.* **92**, 37–40.
12. Kaiser, E. T. and Lo, K. W. (1969) *J. Am. Chem. Soc.* **91**, 4912–4918.
13. Kannan, K. K., Notstrand, B., Fridborg, K., Lövgren, S., Ohlsson, A. & Petef, M. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 51–55.
14. Kannan, K. K., Ramanadham, M. & Jones, T. A. (1984) *Ann. N. Y. Acad. Sci.* **429**, 49–60.
15. Liljas, A., Kannan, K. K., Bergsten, P. C., Waara, I., Strandberg, B., Carlbhom, U., Järup L., Lövgren, S. & Petef, M. (1972) *Nature New Biol.* **235**, 131–137.
16. Notstrand, B., Waara, I. & Kannan, K. K. (1975) in *Isoenzymes, Molecular Structure*, ed. Markert, C. L. (Academic Press, New York), Vol. 1, 575–599.
17. Vinay Kumar, Kannan, K. K. & Chidambaram, R. (1989) *Curr. Sci.* **58**, 344–348.
18. Erickson, A. E. & Liljas, A. (1988) Ph.D. Thesis, Uppsala University.
19. Ramaseshan, S. (ed) (1989) *Curr. Sci.* **58**, 344.
20. Kannan, K. K. (1980a) in *Biophysics and Physiol. of Carbon dioxide*, eds. Bauer, C., Heinz, G. & Bartels, H. (Springer Verlag, Berlin), 184–205.
21. Kannan, K. K. (1980b) in *Biomolecular Structure, Conformation, Function and Evolution: Diffraction and Related Studies*, eds. Srinivasan, R., Subramanian, E. & Yathindra, N. (Pergamon Press, Oxford), Vol. 1 pp. 165–181.
22. Kannan, K. K., Liljas, I., Waara, I., Bergsten, P. C., Fridborg, K., Lövgren, S., Strandberg, B., Bengtson, U., Carlbom, U., Järup, L. & Petef, M. (1971) Coldspring Harbor Symposia on Quant. Biol. **36**, 221–231.
23. Rao, S. T. & Rossmann, M. G. (1973) *J. Mol. Biol.* **76**, 241–256.
24. Creighton, T. E. (1983) in *Proteins, Structures and Molecular Properties*, (Freeman, W. H. and Co., New York), pp. 226.
25. Carlsson, U., Hendersson, L. E., Nyman, P. O. & Samuelson, T. (1974) *FEBS Lett.* **48**, 167–171.
26. Fernley, R. T., Wright, R. D. & Coghlan J. P. (1988) *Biochemistry* **27**, 2815–2820.
27. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963) *J. Mol. Biol.* **7**, 95–99.
28. Erickson, A. E., Jones, T. A. & Liljas, A. (1986) in *Zinc Enzymes*, eds. Bertini, I., Luchinat, C., Maret W. & Zeppezauer (Birkhauses, Bast ib), pp. 317–328.
29. Kannan, K. K., Waara, I., Notstrand, B., Lövgren, S., Borell, A., Fridborg, K. & Petef, M. (1977a) in *Drug Action at the Molecular Level*, ed. Roberts, G. C. K., pp. 73–93.
30. Chakravarty, S., Yadava, V. S., Vinay Kumar & Kannan, K. K. (1985) *Proc. Int. Symp., Biomol., Struct., Interactions, Suppt. J. Biosci.* **8**, 491–498.
31. Erickson, A. E., Jones, T. A. & Liljas, A. (1989) *Proteins* **4**, 274.
32. Kannan, K. K., Vinay Kumar, Sankaran, K., Padma Satyamurthy & Chakravarty, S. (1987) *Curr. Sci.* **56**, 173–174.
33. Khalifha, R. G. (1971) *J. Biol. Chem.* **246**, 2561–2573.
34. Kannan, K. K., Petef, M., Fridborg, K., Hilda, Cid-Dresdner & Lövgren, S. (1977b) *FEBS Lett.* **73**, 115–119.
35. Liang, J. Y. & Lipscomb, W. N. (1989) *Biochemistry* **28**, 9724–9733.
36. Vinay Kumar, Padma Satyamurthy & Kannan, K. K. (1987) *Acta Crystallogr. Supl.* **A43**, 02.1.
37. Kumar, K., King, R. W. & Carey, P. R. (1974) *FEBS Lett.* **48**, 283–287.
38. Gibbons, B. H. & Edsall, J. T. (1964) *J. Biol. Chem.* **239**, 2539–2552.
39. Lindskog, S., Enghers, P., Forsman, C., Ibrahim, S. A., Jonson, B. H., Simonsson, I. & Tibell, L. (1984) *Ann. N. Y. Acad. Sci.* **429**, 61–75.

40. Coleman, J. E. (1984) *Ann. N. Y. Acad. Sci.* **429**, 26–48.
41. Cook, M. C. & Allan, L. C. (1984) *Ann. N. Y. Acad. Sci.* **429**, 84–88.
42. Liang, J. Y. & Lipscomb, W. N. (1987) *Biochemistry* **26**, 5293–5301.
43. Pocker, Y. & Deits, T. L. (1984) *Ann. N. Y. Acad. Sci.* **429**, 76–83.
44. Smith, E. L. (1949) *Proc. Natl. Acad. Sci. U.S.A.* **35**, 80.
45. Davis, R. P. (1959) *J. Am. Chem. Soc.* **81**, 5674–5678.
46. Devoe, H. & Kissinowsky (1961) *J. Am. Chem. Soc.* **83**, 274–279.
47. Woolley, P. (1975) *Nature (London)* **258**, 677–682.
48. Tu, C. & Silverman, D. N. (1977) *J. Biol. Chem.* **252**, 3332–3337.
49. Compbell, I. D., Lindskog, S. & White, A. I. (1974) *J. Mol. Biol.* **90**, 469–489.
50. Merz Jr. K. M., Hoffmann, R. & Dewar, H. J. J. (1989) *J. Am. Chem. Soc.* **111**, 5636–5649.
51. Vedani A., Huhta, D. W. & Jacobes, S. P. (1989) *J. Am. Chem. Soc.* **111**, 4075–4081.
52. Tu, C. K., Silverman, D. N., Forsman, C., Jonsson, B.-H. & Lindskog, S. (1989) *Biochemistry* **28**, 7913–7918.
53. Appleton, D. W. & Sarkar, B. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1686–1690.
54. Pesando, J. M. (1975) *Biochemistry* **14**, 681–688.
55. Martin, R. B. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4346–4347.
56. Coleman, J. E. & Lindskog, S. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 2505.
57. Riepe, M. E. & Wang, J. H. (1968) *J. Biol. Chem.* **243**, 2779–2787.
58. Bertini, I. & Luchinat, C. (1984) *Ann. N. Y. Acad. Sci.* **429**, 89–98.
59. Yeagle, P. L., Lochmuller, C.H. & Henken, R. W. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 454–458.
60. Hendrickson, W. A. & Konnert, J. H. (1981) in *Biomolecular Structure, Conformation Function and Evolution: Diffraction and related studies*, eds. Srinivasan, R., Subramanian, E. & Yathindra, N. (Pergamon Press, Oxford) Vol. 1, pp. 43–57.
61. Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S. H. (1977) *Acta Crystallogr.* **A33**, 800–804.
62. Jones, T. A. (1978) *J. Appl. Crystallogr.* **11**, 268–272.
63. Pullman, A. (1981) *Ann. N. Y. Acad. Sci.* **367**, 340–355.

Variability in Protein Hydration and its Structural Consequences – A Novel Approach involving Water-Mediated Transformations

M. Vijayan and R. Kodandapani

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Biological macromolecules, as indeed living organisms themselves, almost always exist and function in an aqueous environment. The role of water molecules is not confined to providing a suitable medium for biological process; they play a crucial role in the structure, assembly and action of biomolecules. Not surprisingly, therefore, the hydration of proteins and its effect on protein structure and action have received considerable attention over the years (1–6). Furthermore, it has been possible in recent years to define the precise geometrical features of protein–water interactions through the high resolution X-ray analysis of protein crystals (7–14). Such X-ray studies have led to a general picture of water surrounding protein molecules. Each protein molecule is surrounded by a “hydration shell” made up of ordered water molecules which interact directly with protein atoms. A few more ordered water molecules, usually in hydrogen bonded contact with the water molecules in the hydration shell, also normally exist. The water molecules rapidly become disordered as their distance from the protein surface increases and become part of the bulk water that fills crystal interstices. Careful examination of the water structure surrounding the same protein from different species and in different crystal environments appears to indicate that each protein is endowed with its own characteristic hydration shell which is suggested to be, to a great extent, invariant with respect to species variation and environmental effects (10, 12, 15). However, the nature and the extent of the changes that are possible in the hydration shell of a given protein and the effect of these changes on its structure and action are not clearly understood. Indeed, variability in the organisation and the amount of water surrounding protein molecules and its influence on protein structure and function, constitute a problem of fundamental importance. We seek to address this problem through a comparatively novel approach involving water-mediated transformations in protein crystals.

Pioneers in protein crystallography, from its earliest days, have been aware of the transformations in the crystal structure of proteins brought about by changes in the solvent content in the crystals (16, 17). A systematic attempt to exploit such transformations in the crystals of haemoglobin to determine the overall size and shape of the molecule as well as the phase angles of some special X-ray reflections, was made in the late forties and the early fifties (18–20). These experiments were conducted at a time when the first structure solution of proteins was still several years ahead and, understandably, the further ramifications of the results were not explored. We,

however, showed in the mid-eighties that not only haemoglobin but also many other protein crystals undergo reversible transformations with changes in the solvent content in them and that such water-mediated transformations provide a useful approach for exploring the variability in protein hydration and its structural consequences (21, 22).

Preliminary Investigations

Our interest in the variability in protein hydration and its structural consequences arose out of an accidental observation during the X-ray photography of a new form of ribonuclease A (21). The crystals yielded two slightly different diffraction patterns depending upon the relative “dryness” of the environment. Through careful experimentation involving the variation in the relative humidity around the protein crystal, it was established that one of the diffraction patterns corresponded to a normal high humidity form and the other to a low humidity form. Subsequently, a number of typical protein crystals were systematically examined under varying conditions of relative humidity (22). The crystals examined included different crystal forms of bovine ribonuclease A (21, 23, 24), hen egg white lysozyme (25–28), porcine 2Zn insulin (29), porcine 4Zn insulin (30) and the crystals of concanavalin A (31). The examination involved recording X-ray precession photographs from each type of crystals at relative humidities (r.h.) of 100%, 98%, 95%, 93%, 90%, 88%, 84%, 79.5%, 75% and 66%. The r.h.’s of the crystal environment was maintained at the desired values by replacing the mother liquor by appropriate saturated salt solutions (32) in the capillaries containing the crystal. Most crystals substantially or entirely lost their crystallinity at 66% r.h. and detailed examination was confined to the 100%–75% r.h. range.

Of the crystals examined, the new monoclinic form of ribonuclease A grown from an acetone solution in tris buffer (21), the well known monoclinic form of ribonuclease A grown from aqueous ethanol (23), and tetragonal (25), monoclinic (27) and orthorhombic (26) lysozyme underwent one or more reversible transformations, without loss of crystallinity, in the r.h. range 100–75%. The transformations were accompanied by abrupt changes in unit cell volume (figures 1 and 2), diffraction pattern (figures 3 and 4) and the solvent content. The old monoclinic form of ribonuclease A could also be grown from a solution of 2-methylpentane-2,4-diol (MPD) in phosphate buffer (24). This form, 2Zn insulin (29), 4Zn insulin (30) and the crystals of concanavalin A (31) did not transform in the 100–75% r.h. range before the drying process set in. Thus, half of the crystals examined exhibited water-mediated transformation in the 100–75% r.h. range while the other half did not.

The above experiments do not lead to any clear indication of the factors that promote or inhibit water mediated transformation. The crystals examined have varying solvent contents. While the transformation does not appear to depend on the amount of solvent in the crystals it is influenced by the solvent composition. As mentioned earlier, while the well-known monoclinic form of ribonuclease A grown from aqueous ethanol transformed twice in the 100–75% r.h. range, that grown from a MPD solution in phosphate buffer did not transform at all in the same range. Indeed, detailed experiments showed that the presence of MPD has the effect of preventing water-mediated transformation or reducing the r.h. at which it occurs (22). The crystal structure also has a definite influence on the transformation. For example,

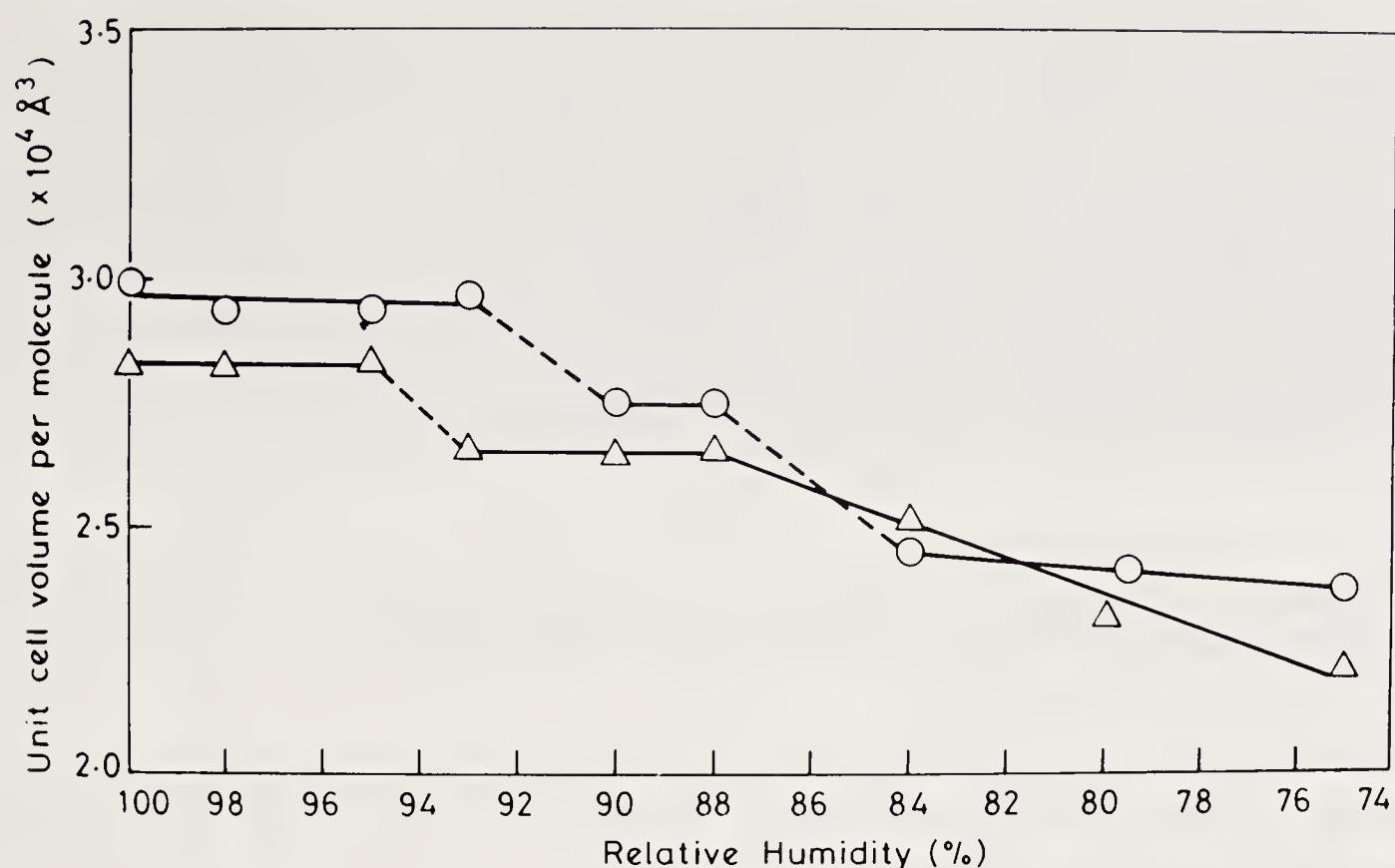


Figure 1. Variation of unit cell volume per molecule as a function of relative humidity in the new crystal form grown from tris-acetone (Δ) and the old crystal form grown from aqueous ethanol (O) of ribonuclease A. Taken from ref. 22.

monoclinic and triclinic lysozyme are grown from the same solvent and have comparable solvent contents. However, one transformed while the other did not.

The results outlined above indicate that water-mediated transformation of protein-crystals is a widespread, though not universal, phenomenon, although the mechanism of and the factors that effect the transformation are not entirely clear. They show that the transformation, when it occurs, cannot be explained simply in terms of the

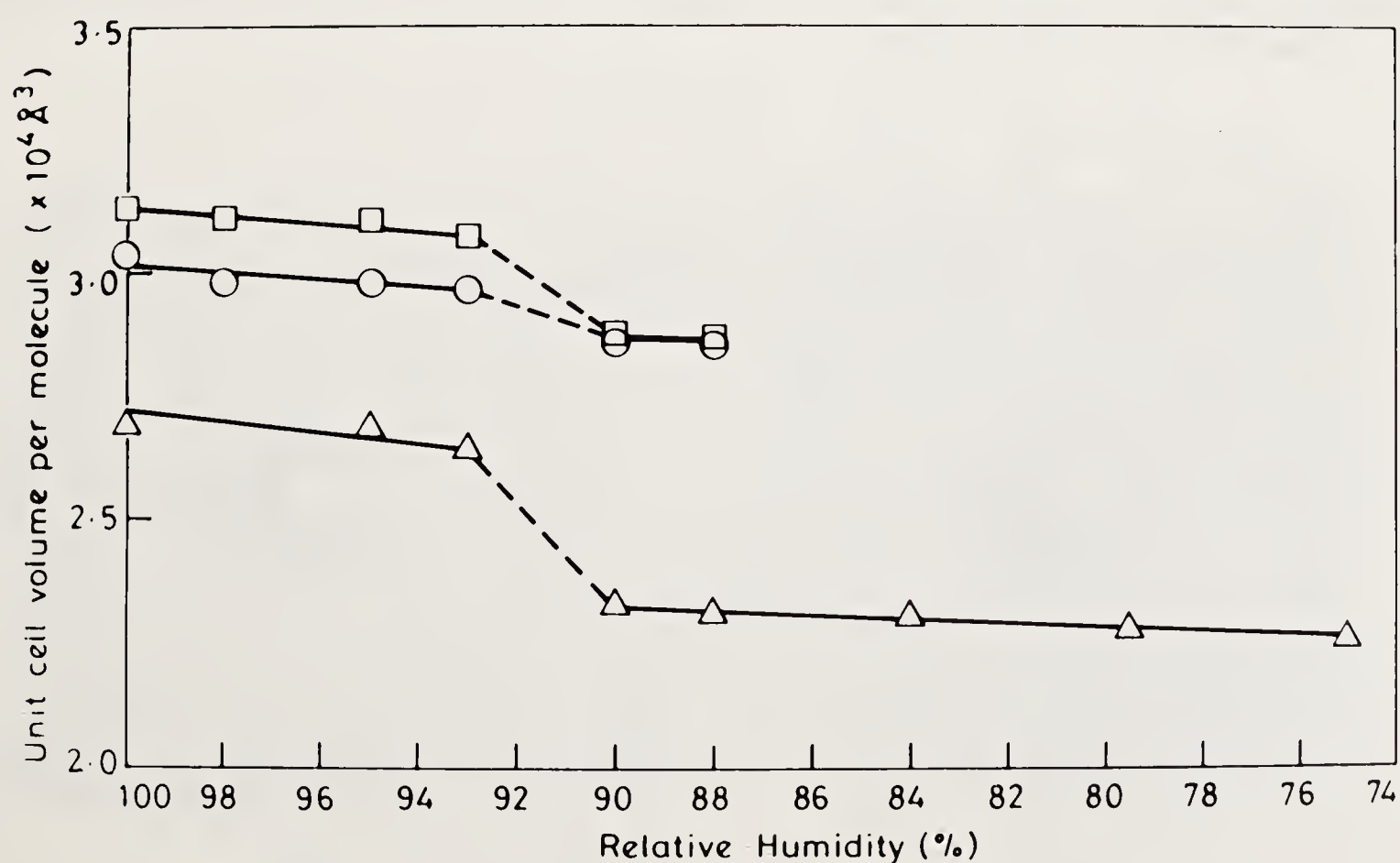


Figure 2. Variation of solvent content as a function of relative humidity in tetragonal (O), orthorhombic (\square) and monoclinic (Δ) lysozyme. Taken from ref. 22.

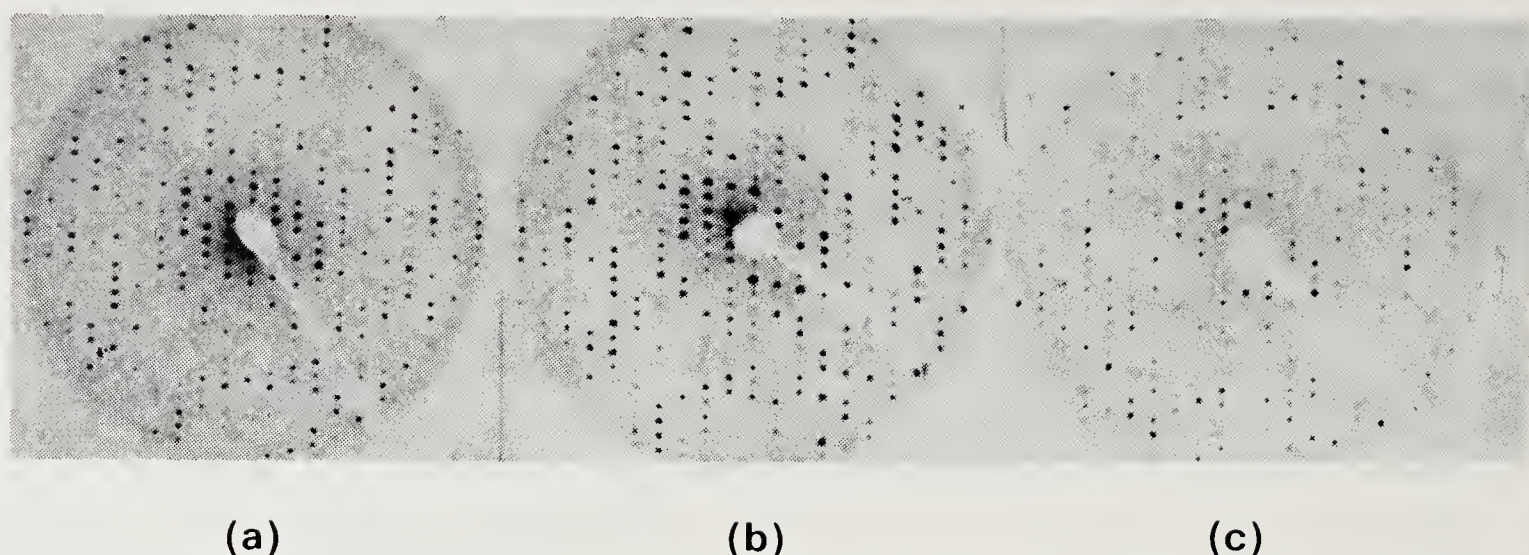


Figure 3. 15° hol precession photographs from the old crystal form of ribonuclease A at (a) 100% (b) 90% and (c) 84% r.h. Taken from ref. 22.

shrinkage of the unit cell due to loss of solvent. *A priori*, it could be accompanied by changes in crystal packing, molecular structure or hydration or, most probably, a combination of the three. It is presumably the bulk water that is removed (or added) during the transformation. The bulk water is contiguous to the water molecules bound to proteins and the changes in the former are most likely to affect the organisation of molecules in the hydration shell and, through the hydration shell, the protein structure itself. Thus water-mediated transformations appeared to provide a novel tool for exploring the variability in protein hydration and its structural consequences. However, the precise changes that occur during the transformation can be elucidated only through the high resolution X-ray analysis of the low humidity forms of protein crystals whose native high humidity structures are already available. As a first step in an effort in this direction we have recently determined the low humidity structure of the well known tetragonal lysozyme (33) and compared it with the refined native structure (34).

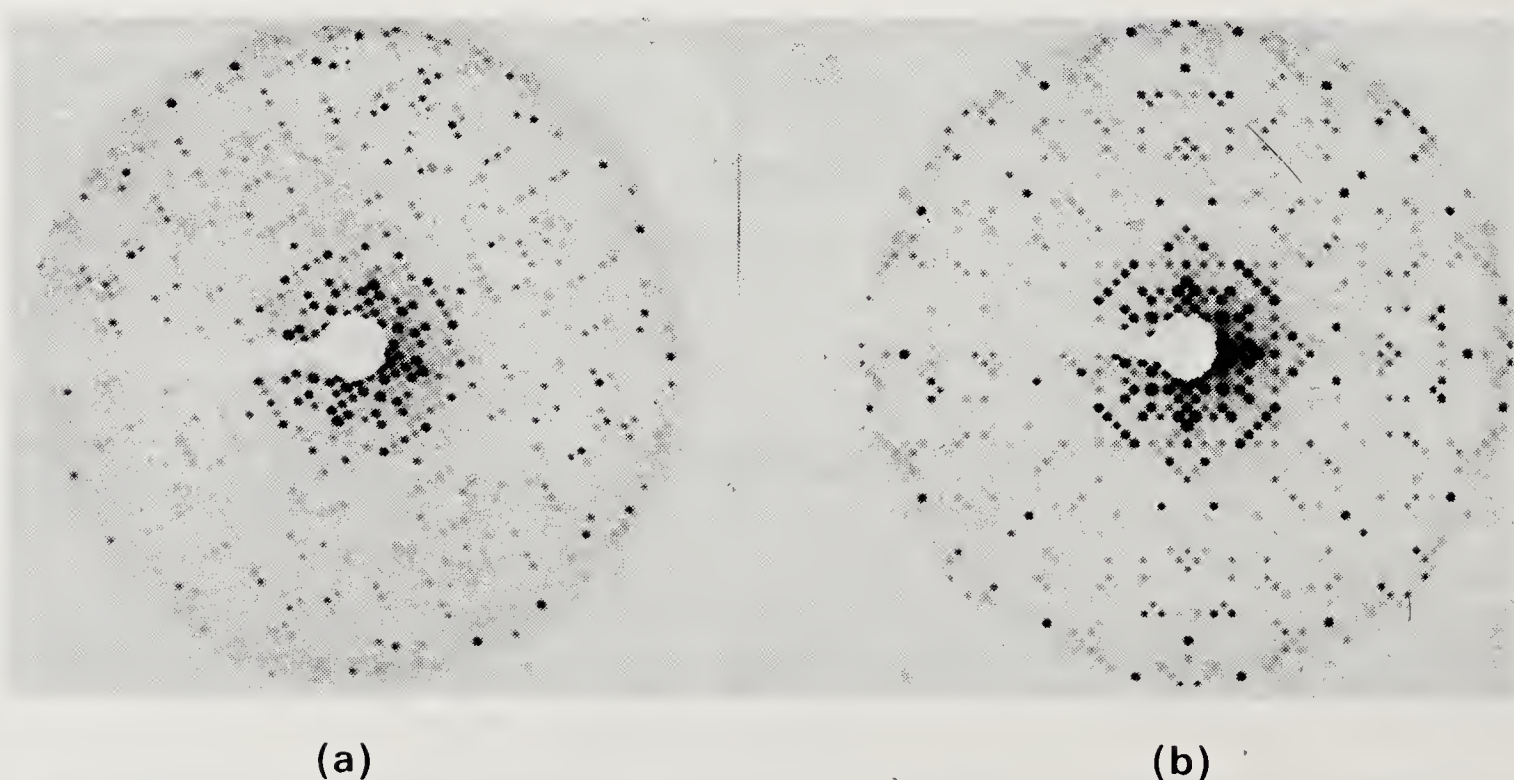


Figure 4. 15° hko precession photographs from tetragonal lysozyme at (a) 100% and (b) 90% r.h. Taken from ref. 21.

Structure of Low Humidity Tetragonal Lysozyme

X-ray Analysis

Tetragonal lysozyme transforms into the low humidity form at around r.h. 90% (21, 22). The X-ray data from the low humidity form ($a = 78.4 \text{ \AA}$; $c = 37.6 \text{ \AA}$; solvent content = 37.8%) were collected upto 2.1 \AA resolution at r.h. 88% and these data were used for structure determination. The final R factor obtained at the completion of the restrained least-squares refinement (35–37) was 0.162 for 6289 observed reflections in the $10\text{--}2.1 \text{ \AA}$ resolution shell (33). The refined structure consists of 1001 protein atoms and 157 water oxygens with an estimated average coordinate error of about 0.15 \AA (38). The criteria employed by Baker and Hubbard (12) were largely used for delineating hydrogen bonds except for a slight increase in the cut off values (2.5 to 2.6 \AA for $\text{H}\dots\text{O}$ and 3.5 to 3.6 \AA for $\text{N}\dots\text{O}$ and $\text{O}\dots\text{O}$ distances) for hydrogen bond distances. The main chain and side chain conformational angles (39, 40) in the structure by and large follow the same patterns as those found in other proteins (41–43).

Molecular Structure

The overall three dimensional structure of hen egg white lysozyme is very well-known (44). However, in the course of the detailed examination of tetragonal low humidity lysozyme, we come across several interesting features of lysozyme geometry, which have not been adequately described before and, therefore, deserve mention here.

It is well known (44) that helices constitute the most important secondary structural feature in lysozyme. As can be seen from the schematic illustration of main chain hydrogen bonds given in figure 5, they include stretches involving residues 5 to 15,

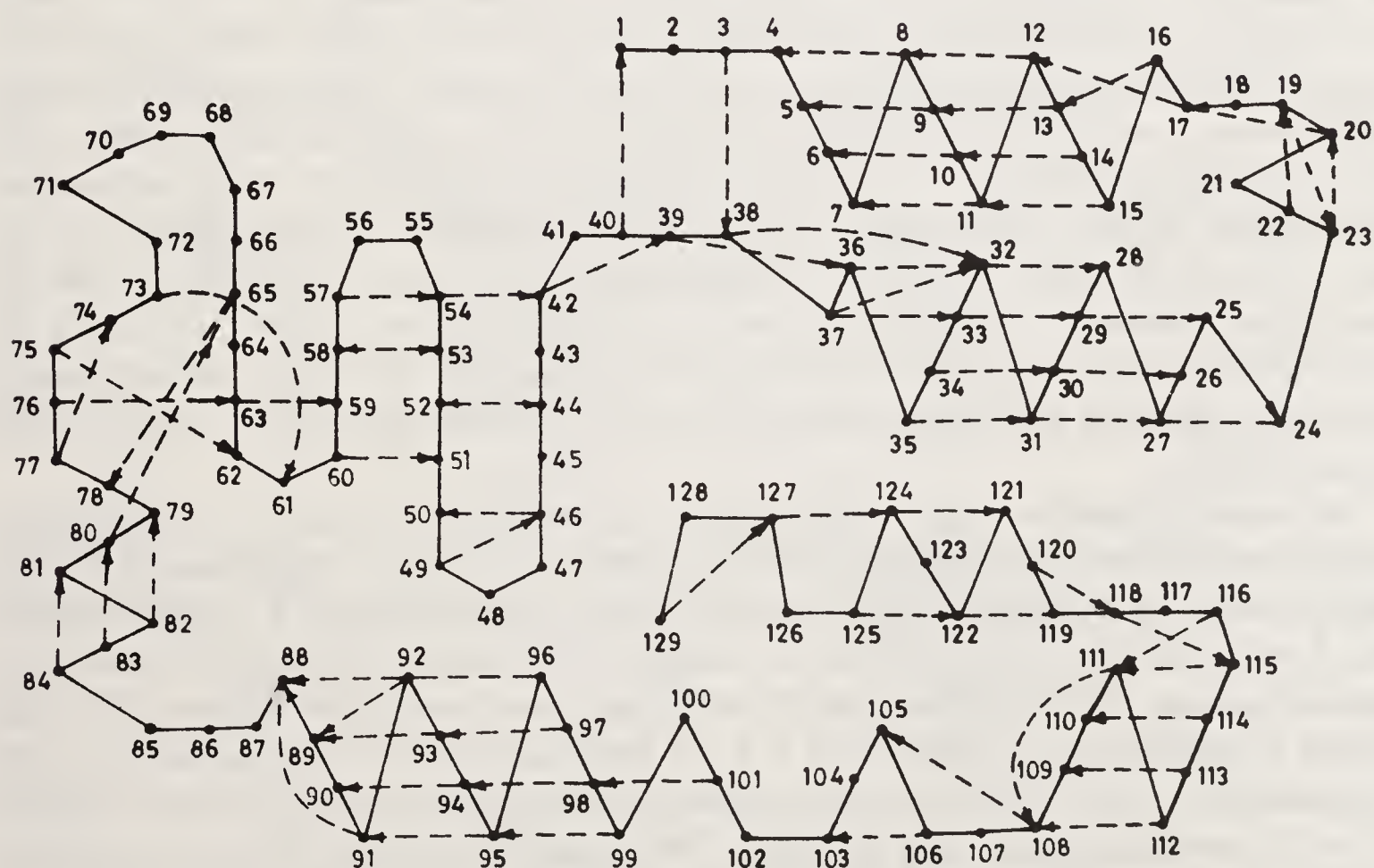


Figure 5. Schematic illustration of hydrogen bonds involving main chain atoms.

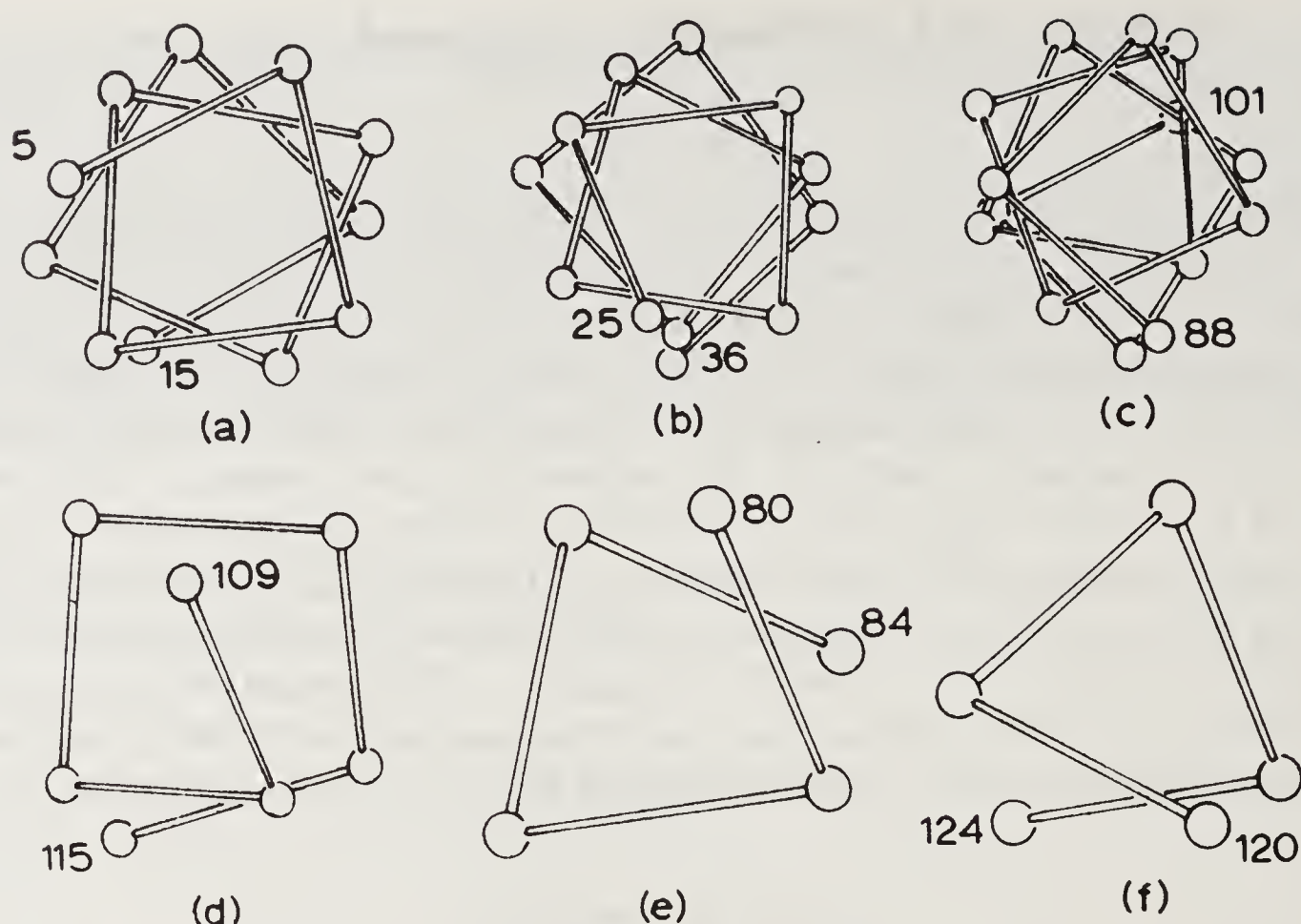


Figure 6. α -Carbon positions in helical stretches, viewed along helix axes. The residues in the helical stretches are. (a) 5–15 (b) 25–36 (c) 88–101 (d) 109–115 (e) 80–84 (f) 120–124.

25 to 36, 88 to 101 and 109 to 115, which are predominantly α -helical in terms of hydrogen bonding, and the 3_{10} stretches 80 to 84 and 120 to 124. These helices are illustrated in figure 6. Stretches 5 to 15, 25 to 36 and 88 to 101, particularly the first two, are good α -helices with average unit twist (θ) and rise (t) per residue (45, 46) close to those for a classical α -helix. The 109–115 stretch (mean $\theta = 97(31)^\circ$, mean $t = 1.5(7) \text{ \AA}$), however, exhibits considerable departure from ideal α -helical geometry. In fact, if residue 109 is omitted from consideration, the mean θ and t for the remaining residues work out to be $85(9)^\circ$ and $1.3(7) \text{ \AA}$, values close to those for a π -helix (47). The stretch 80 to 84 (mean $\theta = 110(20)^\circ$ and mean $t = 1.8(2) \text{ \AA}$) have helical parameters intermediate between those for 3_{10} and α -helices although it is described as a 3_{10} helix on the basis of the hydrogen bonding pattern. The same is true of the 120–124 stretch also (mean $\theta = 112(22)^\circ$ and mean $t = 2.0(2) \text{ \AA}$). In fact, in both the cases, the first and the last sets of parameters along the polypeptide chain have values close to those for a classical 3_{10} helix, whereas those in the middle have values close to those for an α -helix.

The major β -structure region made up of residues 42 to 63 (figure 7) contains two features which have not been reported in lysozyme so far. They are a classical β -bulge (48, 49) involving residues 49 and 50, and a 5 residue (residues 59 to 63) turn stabilized by a $5 \rightarrow 1$ hydrogen bond. Elsewhere, residues 19–22 and residues 20–23 present an interesting case, illustrated in figure 8, involving two consecutive β -turns (50). This feature is additionally stabilized by a $1 \rightarrow 5$ hydrogen bond. The consecutive turns are preceded by the 17–20 turn, thus making the segment between the two α -helices in the N-terminal region rich in turns. The segment between the two C-terminal α -helices also is rich in turns with two of them (103–106 and 105–108) occurring in

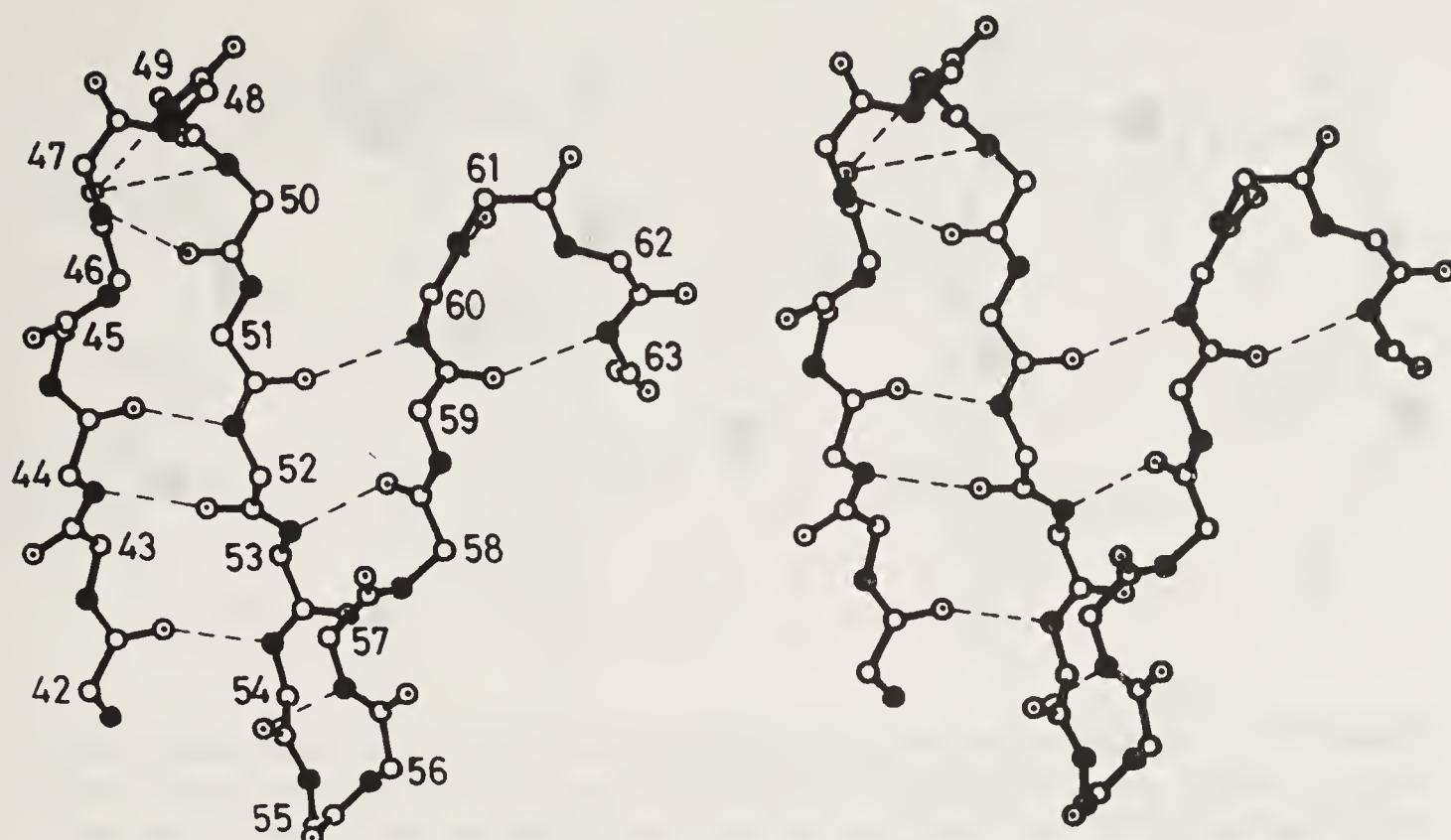


Figure 7. Stereo view of the peptide chain in the major β -structure region. In this and many subsequent figures, side chains are omitted entirely or partially for clarity. In all the figures carbon atoms are represented by open circles, nitrogens by shaded circles and oxygens by dotted circles. Unless otherwise indicated, the residue numbers are given close to the appropriate C_α positions.

quick succession. Two inverse type γ -turns (51, 52) involving 118–120 and 127–129 could also be identified in the structure. It is interesting to note that out of the 75 main chain–main chain hydrogen bonds in the structure, as many as 68 occur in different secondary structural regions.

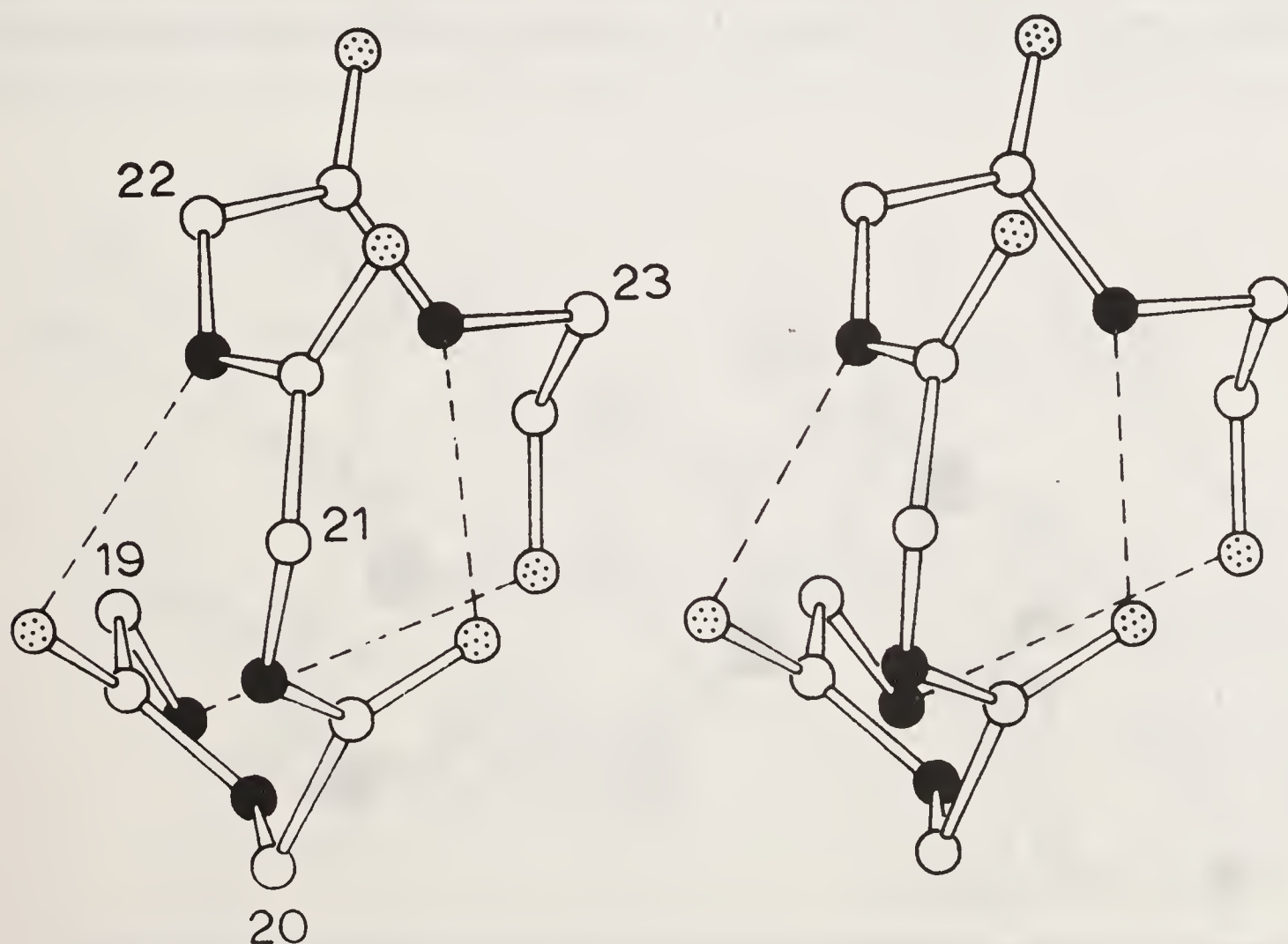


Figure 8. Stereo view of consecutive β -turns between residues 19 and 23.

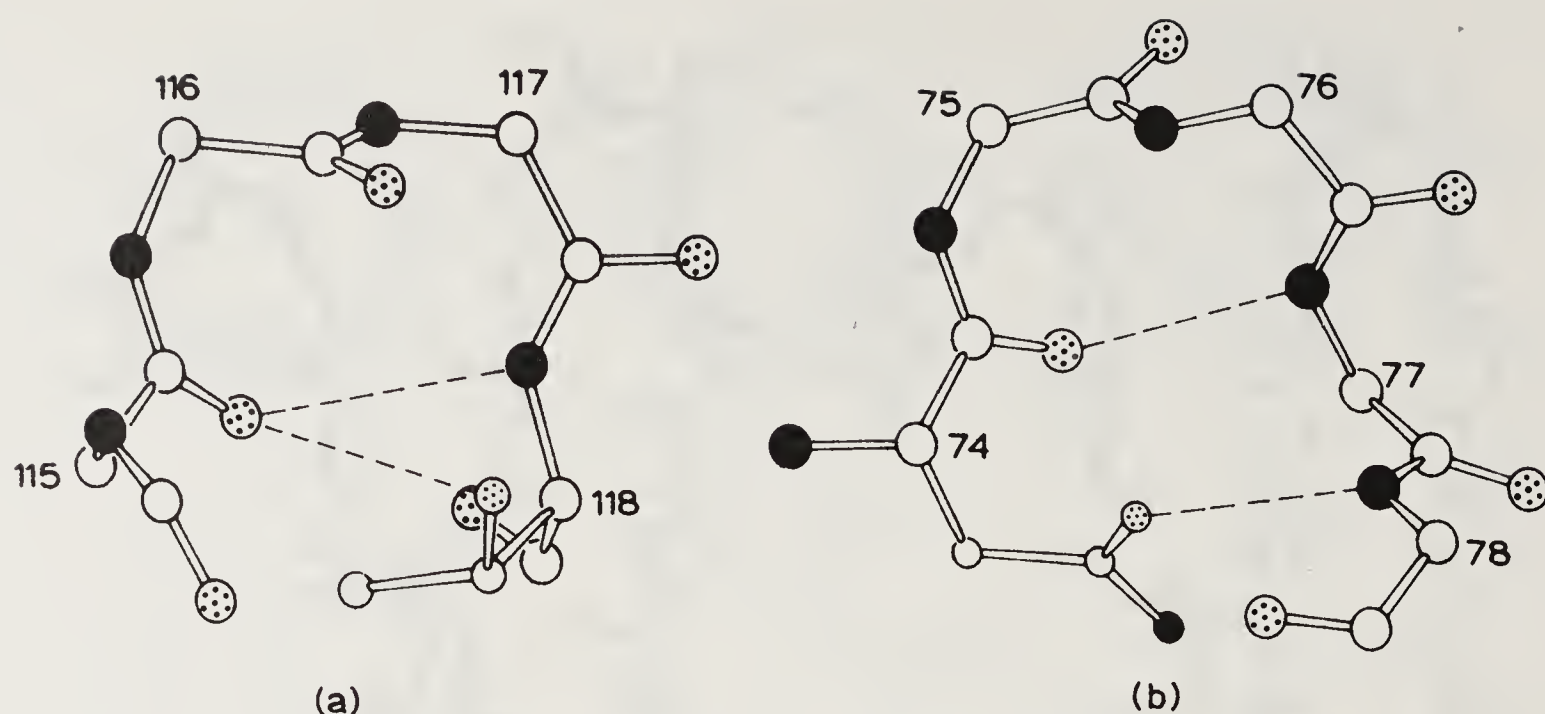


Figure 9. Examples for the further stabilization, a β -turn by a main chain–side chain hydrogen bond with (a) a side chain atom of the 4th residue as donor and the CO group of the first residue as an acceptor and (b) the NH group of the 5th residue as the donor and a side chain oxygen of the first residue as the acceptor. Taken from ref. 33.

The structure presents interesting examples of the role of interactions involving side chains in additionally stabilizing secondary structural features. In particular, reverse turns are additionally stabilized by main chain–side chain hydrogen bonds. Eight such interactions are associated with β -turns, a couple of which are illustrated in figure 9. The main chain–side chain hydrogen bond associated with the lone α -turn in the structure is shown in figure 10. Another interesting main chain–side chain hydrogen bond, illustrated in figure 11, is involved in additionally stabilizing the

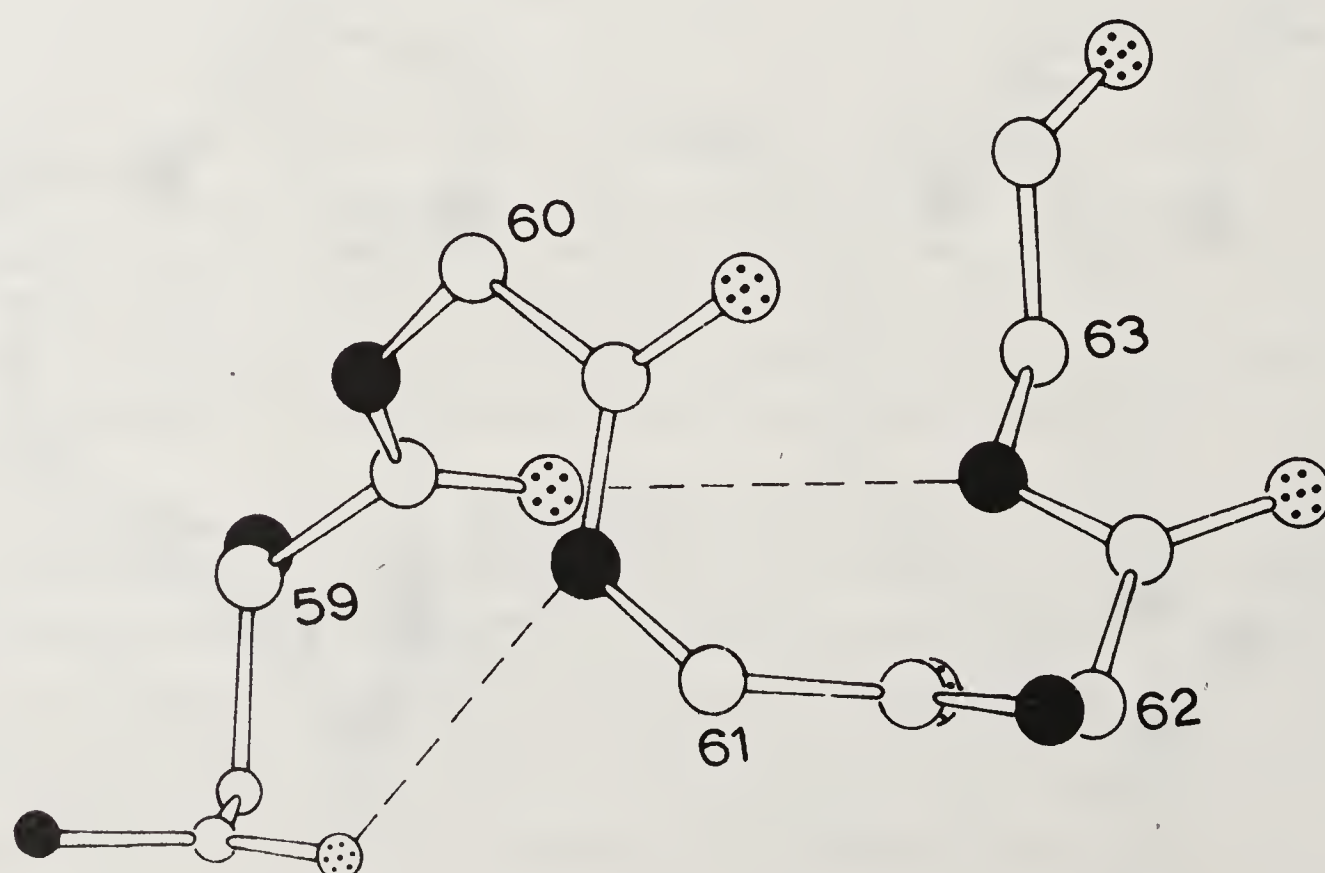


Figure 10. Further stabilization of the 59–63 α -turn by a hydrogen bond between 61 NH and 59 O δ 1.

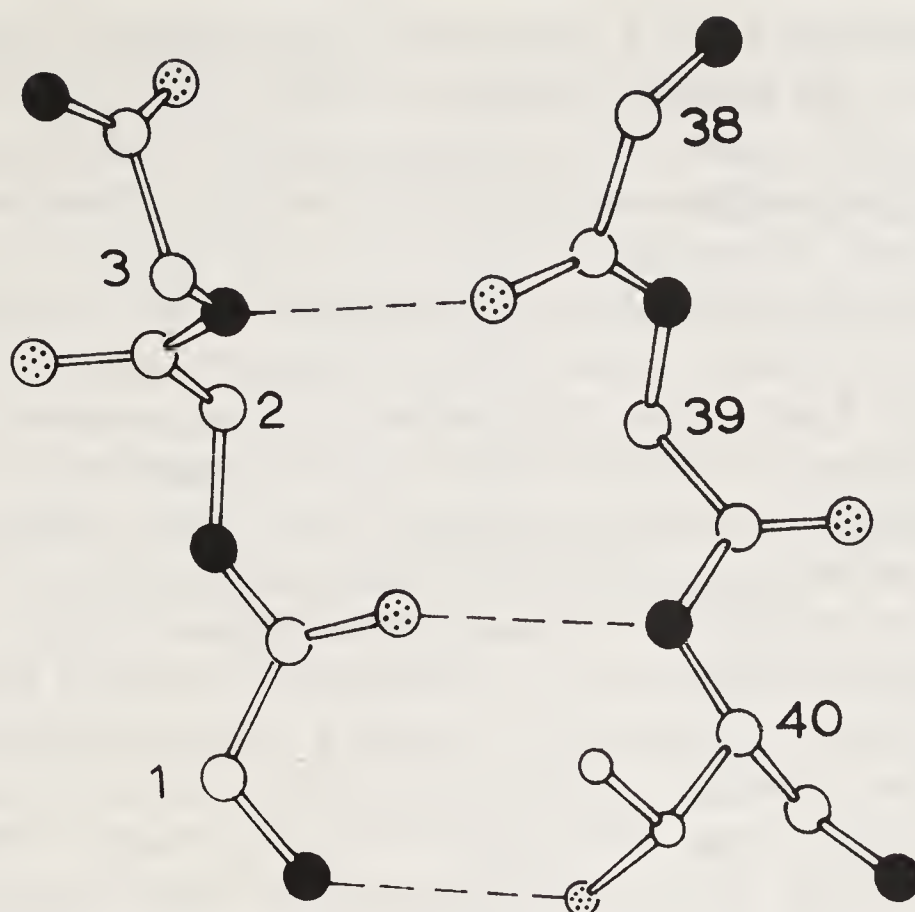


Figure 11. Further stabilization of the short stretch of antiparallel β -strands involving residues 1 to 3 and 38 to 40, by a side chain–main chain hydrogen bond.

minor β -structure region. The main β -structure region is rich in side chain–side chain hydrogen bonds. These hydrogen bonds, shown in figure 12, appear to lend additional stability to the β -structure.

Water Structure

The focus of our attention has been the variability in protein hydration and its consequences and, therefore, the water structure in the low humidity form was

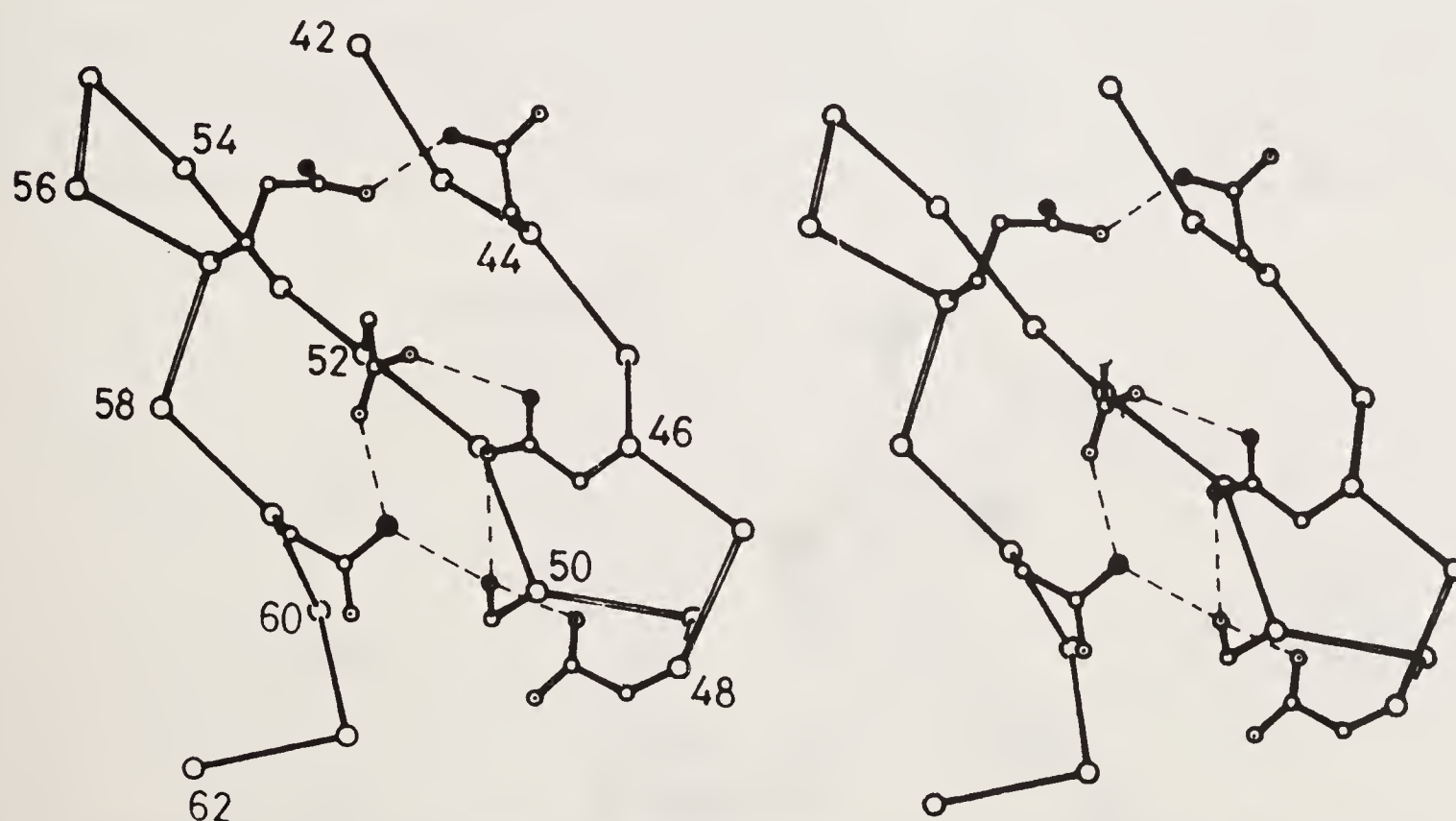


Figure 12. Stereo view of the side chain–side chain interactions in the main β -structure region.

examined in considerable detail. A total of 157 water molecules were identified in the structure. Purely on the distance criterion ($< 3.6 \text{ \AA}$), 107 of them interact with the nitrogen or the oxygen atoms in the protein molecule. 16 of these water molecules are involved in bridging neighbouring protein molecules. Of these, two bridges involve twofold related atoms. Taking these factors into consideration, it is easily seen that each protein molecule is surrounded by and interacts with 121 water molecules. 64 of these have only one interaction with a given protein molecule. 31 interact with two protein atoms and 13 with three. The number of water molecules with 4, 5 and 6 interactions with the protein molecule are 9, 3 and 1 respectively. In terms of the role of water in stabilization of protein structure, it is the water molecules with multiple interactions with the protein molecule that are of obvious interest. Each such water molecule connects or bridges two or more protein atoms. When a water molecule is involved in two protein interactions, it is involved in one water bridge. Considering any pair from a cluster of protein atoms around a water molecule as contributing to a bridge, the number of bridges mediated by a water molecule increases rapidly with the number of protein atoms with which it interacts. Such bridges predominately involve local interactions. Of the 169 intramolecular water bridges in the structure, nearly 68% (115 in number) involve protein atoms separated by 5 residues or less.

Many of the local water bridges appear to correspond to what may be described as favourable sites of hydration. For instance, 114 Arg and 128 Arg interact specifically (53) with two water molecules each, as illustrated in figure 13. One of the two possible specific interactions is realised in three other arginine residues. Several side chain carboxylate and amide groups are also involved in similar specific interactions. One such interaction is shown in figure 14. Thus, side chains with multiple hydrogen bonding centres appear to be favourable sites of hydration. There are three instances where the main chain NH and CO groups of the same residue interact with a water molecule. There are eight instances where a water molecule is held together by the oxygen or the nitrogen of a short hydrophilic side chain and the main chain NH or

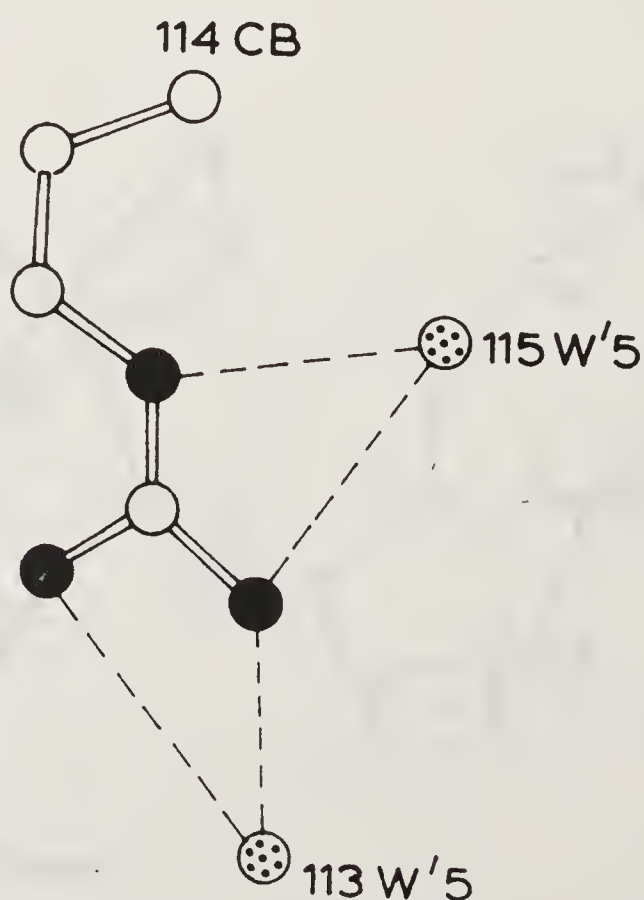


Figure 13. Specific interactions of the guanidyl group of 114 Arg with water molecules. Taken from ref. 33.

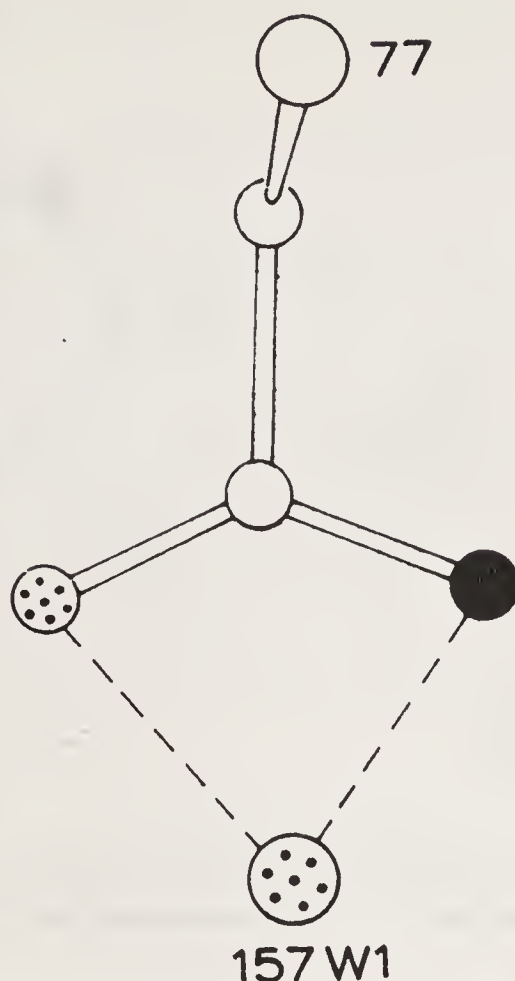


Figure 14. An example of a specific interaction between a side chain amide group and a water molecule.

CO group of the same residue. Two such hydration sites are illustrated in figure 15. As illustrated in figure 16, favourable sites of hydration could also be provided by such side chains and the NH or the CO group of the nearest or the next nearest residue. These side chains often belong to seryl, threonyl, aspareryl or aspartic acid residues.

Water bridges involving side chain atoms are dominated by residues with short hydrophilic side chains. The guanidyl group of the long arginyl side chain is involved in several water bridges, especially between atoms separated widely along the polypeptide chain. In general, the number of water bridges decreases as the distance along the polypeptide chain between the participating atoms increases. This is particularly so of water bridges involving exclusively main chain atoms. About half the bridges that connect atoms separated by five residues or less are of this type

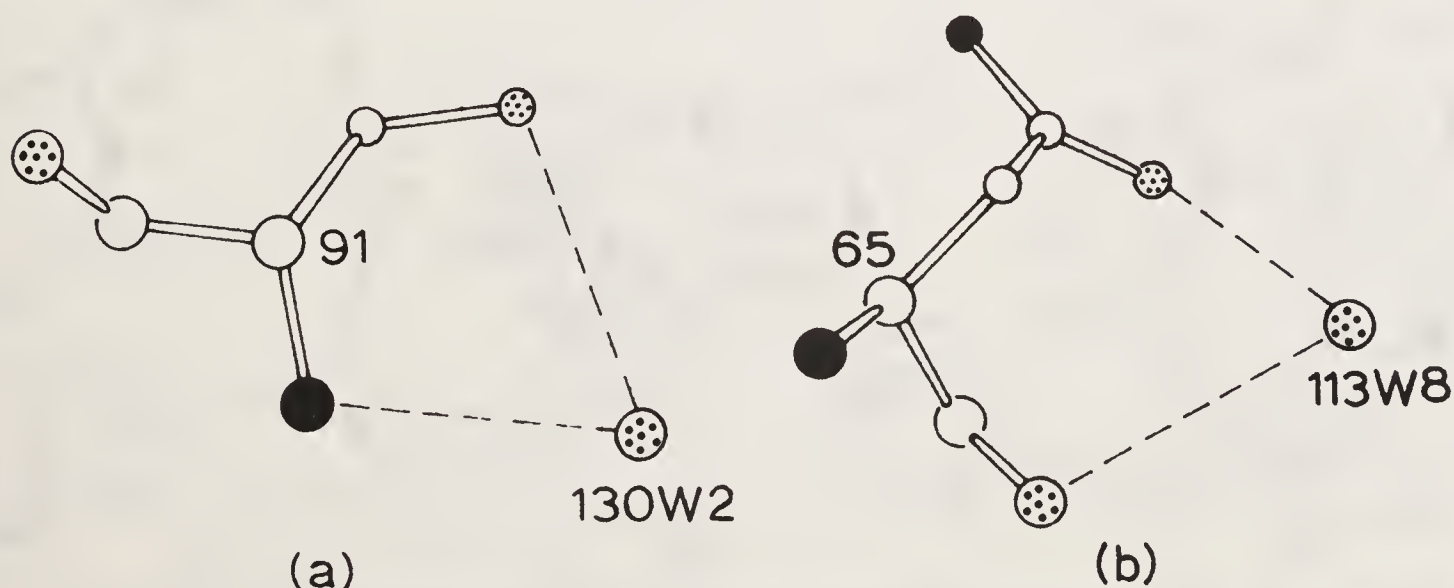


Figure 15. Examples of water bridges between a side chain oxygen and (a) the main chain NH and (b) the main chain CO, of the same residue. Taken from ref. 33.

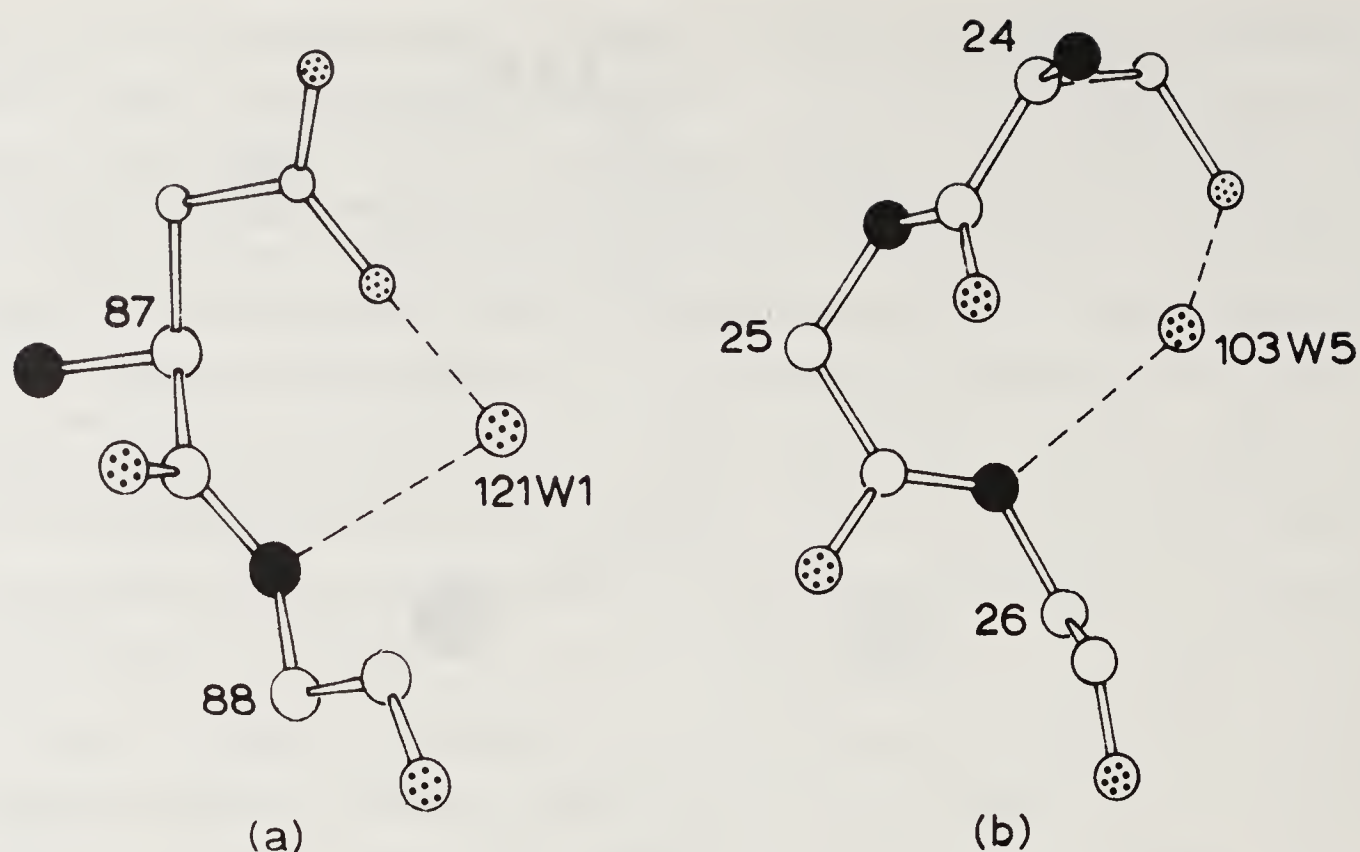


Figure 16. Examples of water sites between a side chain and the NH group of (a) the nearest residue and (b) the next nearest residue. Taken from ref. 33.

whereas bridges involving exclusively main chain atoms constitute only a small fraction of those between atoms separated by a larger number of residues.

The effect of bound water molecules on secondary structural features is an interesting topic (9, 10, 11). In the present structure, NH and CO groups involved in hydrogen bonds that stabilize helices are not, by and large, hydrated. The appropriate hydrogen bonds are not disrupted even in the few cases where they are. The effect of hydration on helices could therefore be considered as marginal, if any. The same is true about the main β -structure region also. However, a well defined water molecule which interacts with 1 Lys N, 3 Asn O $^{\delta 1}$ and 41 Gln N $^{\epsilon 2}$ helps in holding together the two strands in the minor β -region. Not surprisingly, the residues involved in turns are heavily hydrated. As can be seen from a few examples illustrated in figure 17, water bridges often appear to lend additional stability to the turns. These bridges could involve not only the first and the fourth residues in the turn but also those preceding and succeeding the turn.

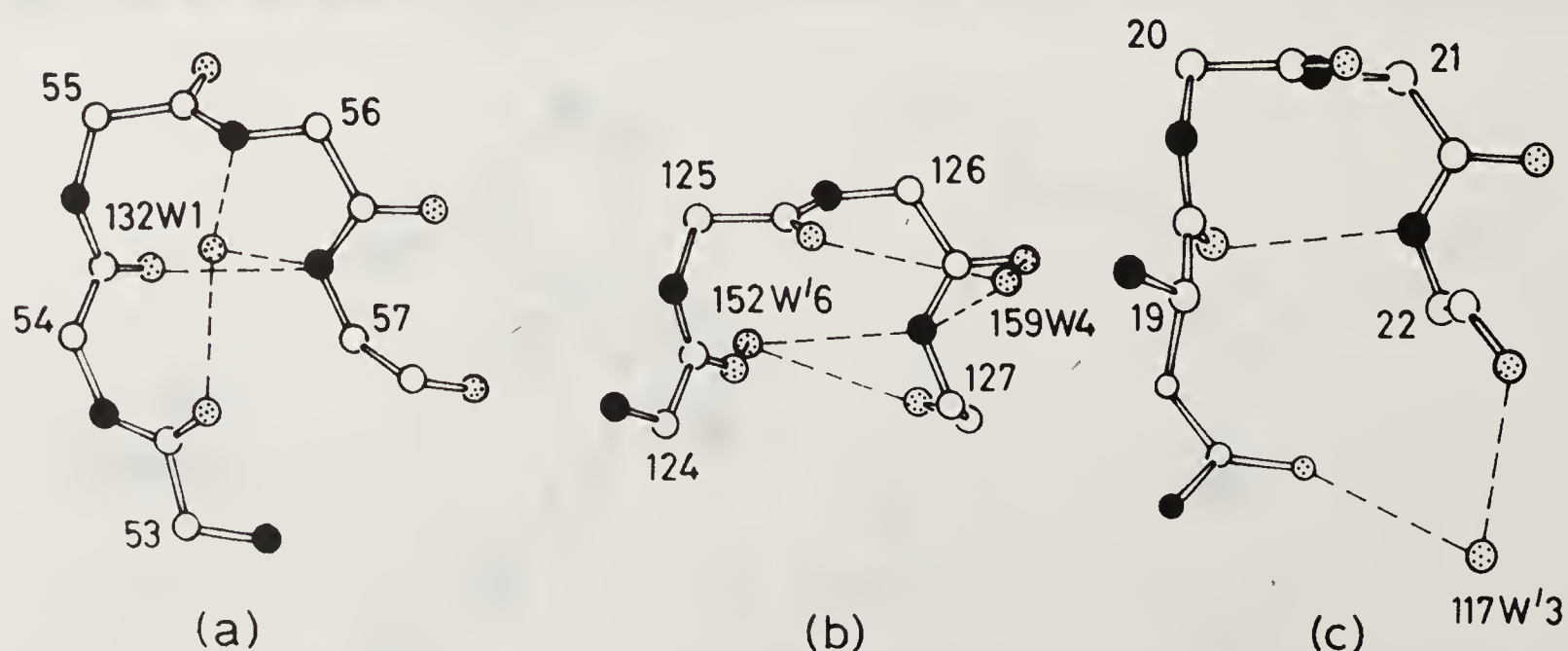


Figure 17. A few examples of water molecules lending additional stability to β -turns. Taken from ref. 33.

Table 1. List of water molecules with $B < 30$ and having three or more interactions with protein atoms.

Water	B	Interacting atoms
W5' 113	13 Å ²	22 Gly O, 24 Ser N, 24 Ser O ^γ , 27 Asn N ^{δ2} , W3' 110.
W3 115	15	1 Lys N, 39 Asn O ^{δ1} , 41 Gln N ^{ε2} .
W4 115	29	118 Thr O ^{γ1} , 119 Asp O, 123 Trp N ^{ε1} .
W1 123	17	41 Gln O, 43 Thr N, 43 Thr O ^{γ1} .
W1' 123	17	53 Tyr OH, 68 Arg N ^ε , 68 Arg NH1.
W2 124	20	112 Arg O, 113 Asn O, 114 Arg N, 116 Lys O.
W1 126	24	83 Leu O, 85 Ser O, 91 Ser O ^γ , W2 130.
W2 130	10	82 Ala O, 83 Leu O, 87 Asp O, 91 Ser N, 91 Ser O ^γ , W1 126.
W1 131	27	87 Asp O, 87 Asp O ^{δ1} , 89 Thr N, 90 Ala N, W1 133.
W1 132	13	53 Thr O, 56 Leu N, 57 Gln N, 91 Ser O ^γ .
W2 142	21	49 Gly O, 51 Thr N, 51 Thr O ^{γ1} , 66 Asp O ^{δ2} , 68 Arg N ^ε , 69 Thr O ^{γ1} .
W9 145	23	98 Ile O, 102 Gly N, 103 Asn N, 104 Gly O, W7 148.
W3 147	29	21 Arg NH2, 99 Val O, 102 Gly N, 103 Asn N, 104 Gly N.
W1 149	22	64 Cys O, 65 Asn N ^{δ2} , 66 Asp O ^{δ1} , 72 Ser O ^γ , W2 152.
W2 152	13	65 Asn N ^{δ2} , 68 Arg N, 69 Thr N, 69 Thr O, W1 149.

Water molecules with $B < 30$ and having three or more interactions with protein atoms are listed in table 1. All the three water molecules described as trapped inside the enzyme molecule in the native structure (44) are present with comparatively low B values in the low humidity form. The first of these (OW2 130) interacts with 82 Ala O, 83 Leu O, 87 Asp O, 91 Ser N, 91 Ser O^γ and a second water molecule (OW1 126) (figure 18). The second water molecule in turn interacts with 83 Leu O, 85 Ser O and 91 Ser O^γ. Thus the water molecules between them connect the 80–84 3_{10} helix with the 88–101 α -helix. The third water molecule (OW1 132) mentioned by Imato *et al.* (44) interacts with 53 Tyr O, 56 Leu N, 57 Gln N and 91 Ser O^γ, and thus bridges the main β -structure and the 88–101 α -helix. There are also other well defined water

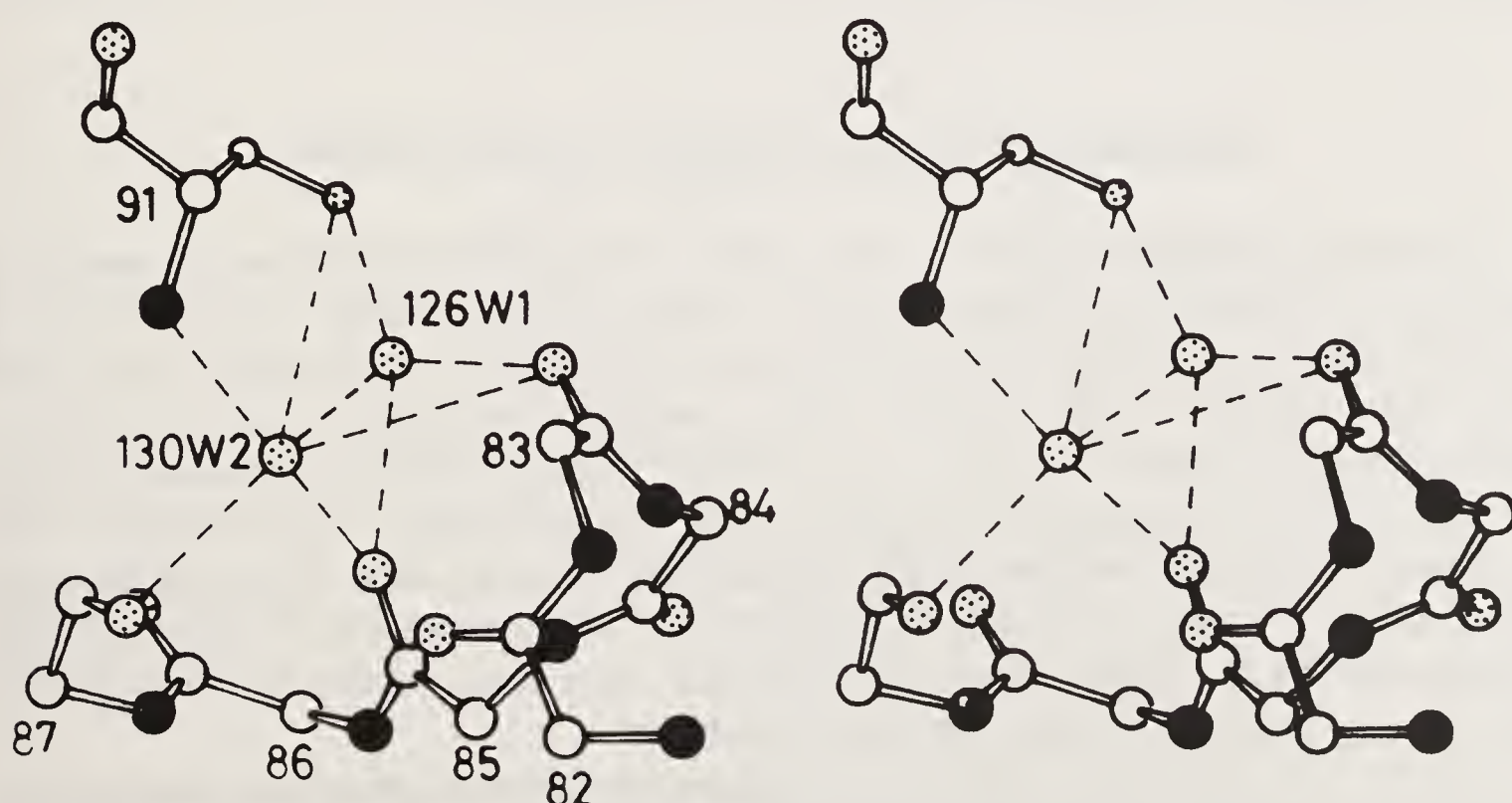


Figure 18. Stereo view of interactions involving two internal water molecules which connect 80–84 and 88–101 helices.

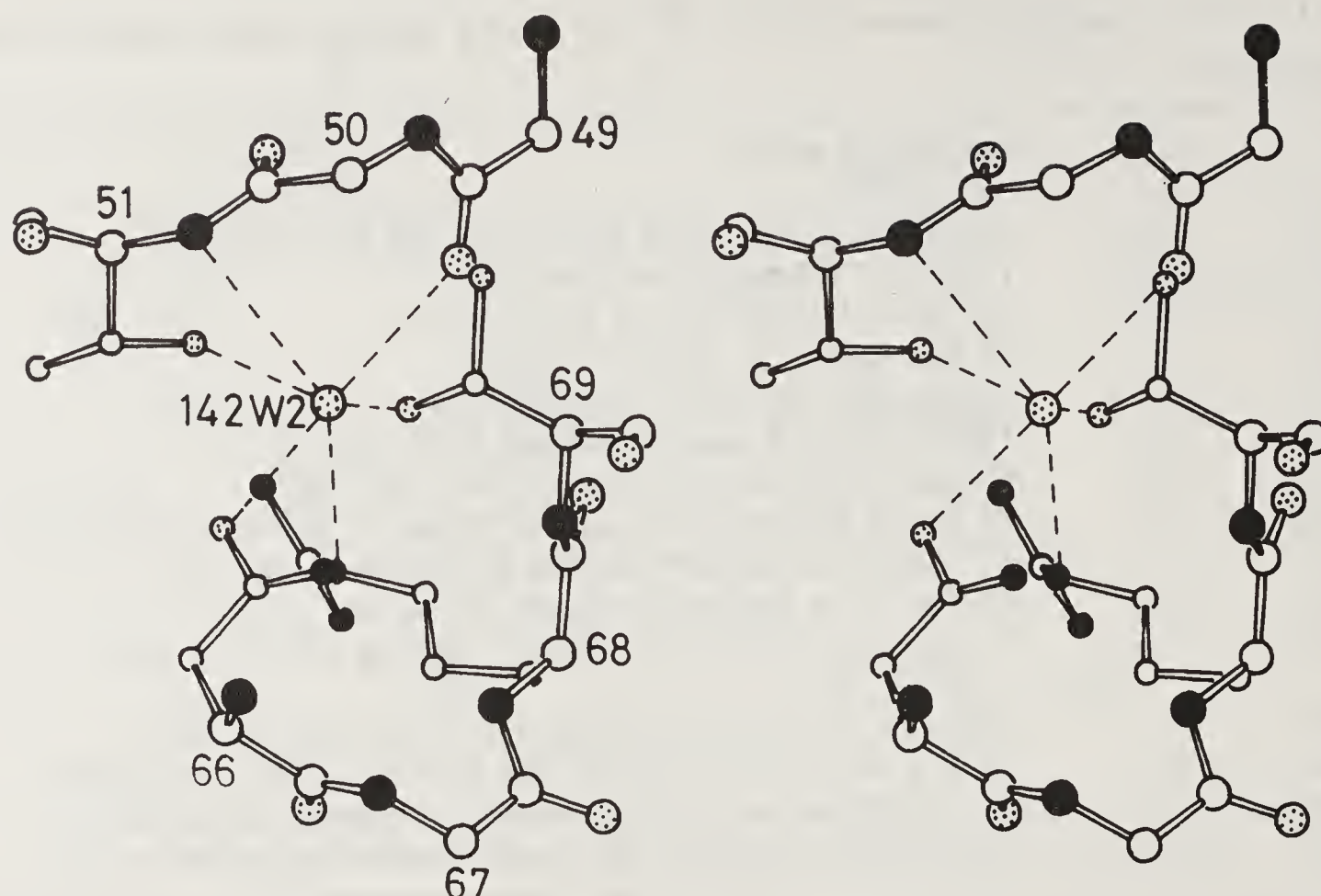


Figure 19. Stereo view of interactions involving a water molecule which connects the β -structure and the loop regions.

molecules which connect different regions of the protein molecule. For example, OW2 142 (figure 19) and OW1' 123 connect the β -structure and the loop. There are other such water molecules which appear to stabilize the local structure. For example, OW1 149 and OW2 152 appear to lend additional stability to the loop structure. OW5' 113 lends additional stability to the region at the start of the 25–35 α -helix. The same is true in relation to the 88–101 α -helix. OW9 145 and OW3 147 interact with atoms at the carboxyl end of this helix and just beyond. Four atoms at the carboxyl end of the 109–115 helix are attached to OW2 124. Indeed, the termini of α -helices appear to be favourable locations for anchoring water molecules.

Structural Changes during Transformation

The water-mediated transformation, which is completely reversible, is caused by the gentle withdrawal of a small amount of water, corresponding to about 25 out of a total of close to 400 water molecules per protein molecule in the native form, from the solvent regions in the crystal. This is perhaps the mildest way to cause a transformation in protein crystals. The changes that accompany the transformation are thus likely to correspond to intrinsic structural variability. These changes were elucidated by carefully comparing the low humidity structure with the native structure, the atomic coordinates and thermal parameters of which were kindly made available by Professor D. C. Phillips. Both the structures have been refined to comparable *R* factors using data upto nearly the same resolution.

In order to facilitate the comparison between the structures, the α -carbon positions in the low humidity form were superposed on those in the native crystals using a method developed by Rossmann and Argos (54), and adapted to the Dec-1090

computer by Dr. M. R. N. Murthy. The transformation matrix for doing so worked out to be

$$\begin{array}{ccc} 0.9999 & -0.0129 & -0.0046 \\ 0.0129 & 0.9999 & -0.0010 \\ 0.0046 & 0.0009 & 1.0000 \end{array}$$

when both the coordinate sets were in Å, with translations of 0.64, 0.33 and 0.03 Å along a , b and c . Thus, the molecule as a whole undergoes a slight rotation and significant translation during the transformation. The translational shift is anisotropic with respect to the unit cell edges and cannot be entirely explained in terms of the shrinkage of the cell. Using the method developed by Richards and co-workers (55), no hinge movement was observed during the water mediated transformation.

Differences in Molecular Geometry

The molecular structure in the low humidity form remains essentially the same as that in the native crystals. Small, though significant, differences however exist. In order to explore these differences, the molecules in the low humidity form and the native form were superposed using the orientation matrix and translation based on α -carbon positions. 10 out of a total of 516 main chain atoms shift by more than 0.5 Å during the transformation. The r.m.s. change for all main chain atoms is as low as 0.21 Å. The r.m.s. change in ϕ and ψ angles is 7° . The numbers of ϕ and ψ angles which shift by 10° or more are 18 and 25 respectively. There are six residues in which both ϕ as well as ψ change by 10° or more. Understandably, the shifts in the positions of the side chain atoms are larger. 113 out of a total of 484 side chain atoms shift by 0.5 Å or more, the r.m.s. shift in all side chain atoms being 0.72 Å. Correspondingly the changes observed in side chain torsion angles are higher than those in the main chain torsion angles.

Minor differences exist between the two structures in hydrogen bonding, but most of them do not appear to be particularly significant. However, absence of change in hydrogen bonding does not always signify absence of structural change. The movements of 33 Lys and 37 Asn, illustrated in figure 20, provide a good example

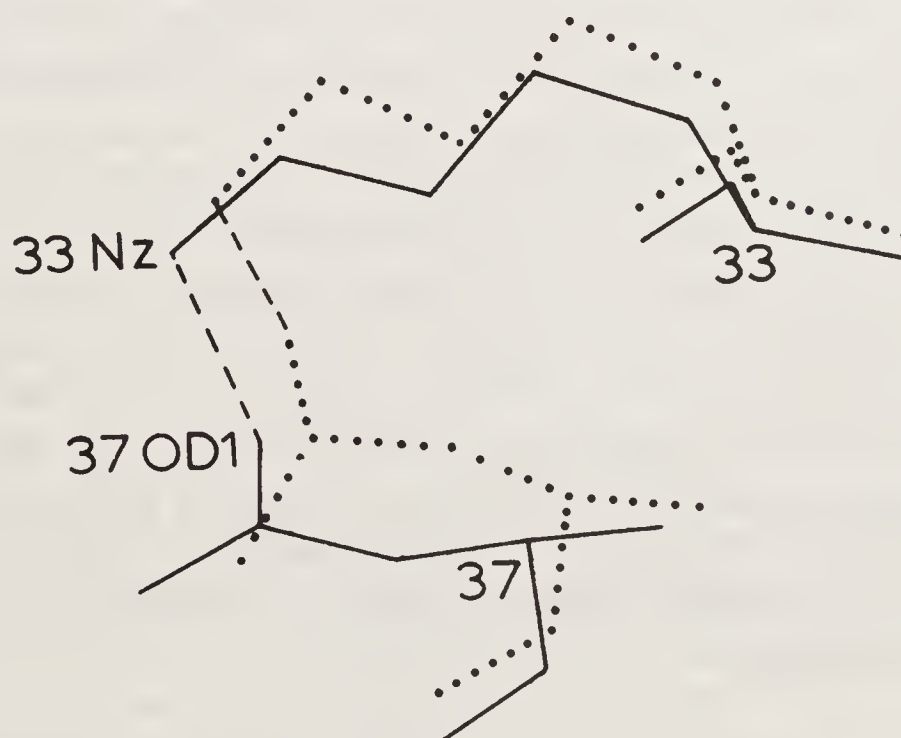


Figure 20. Illustration of the concerted movement of Lys 33 and Asn 37. In this and the next figure, the residues in the native and the low humidity forms are shown by unbroken and dotted lines respectively. Taken from ref. 33.

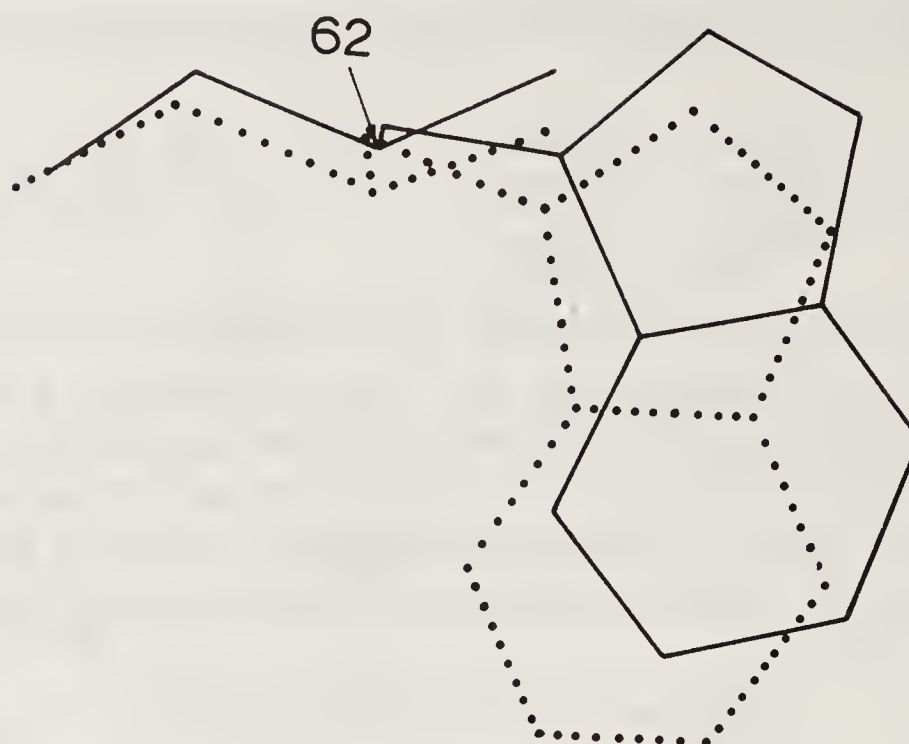


Figure 21. Illustration of the movement of Trp 62 during water mediated transformation. Taken from ref. 33.

for this observation. The side chain amino nitrogen of the lysyl residue is hydrogen bonded to the side chain amide oxygen of the asparagyl residue. The two atoms move by more than 0.5 \AA and 1 \AA respectively during the transformation. However, the two side chains move in a concerted fashion such that the hydrogen bond is retained. 125 Arg NH₂...119 Asp O ^{δ 2}, 112 Arg N ^{ϵ} ...106 Asn O, 72 Ser O ^{γ} ...64 Cys O and 91 Ser O ^{γ} ...83 Leu O are among the other hydrogen bonds which are retained in a similar fashion despite atomic movements of 0.5 \AA or above.

The regions of the molecule that move during water-mediated transformation are substantially the same as those which do during inhibitor binding (44, 56, 57, 58, 59). The movement of 62 Trp, illustrated in figure 21, is particularly noteworthy. The side chain of this residue moves during the transformation by as much as 0.75 \AA on an average with a maximum atomic movement of 1.26 \AA . The loop region also moves substantially during the transformation as it does during inhibitor binding. Indeed, half the main chain atoms that move by 0.5 \AA or more belong to this region. Apart from 62 Trp, 37 Asn and 44 Asn are the active site residues which move substantially during the transformation. So far it has not been possible to correlate in any simple manner the movements during inhibitor binding with those during the transformation. However, the commonality of the residues that move during the two situations emphasizes the importance of the flexibility of certain regions of the enzyme molecule in its action.

The only other study in which tetragonal lysozyme has been studied under a condition other than that obtained in the native crystals is through the application of a hydrostatic pressure of 1000 atmospheres (55, 60). The changes in molecular geometry caused by high pressure, with r.m.s. shifts of 0.12 \AA and 0.20 \AA for main chain atoms and all atoms respectively, are much smaller than those caused by the water mediated transformation. Thus the effect of a gentle withdrawal of a few water molecules from the surrounding medium has a much higher effect on protein structure than the drastic environmental change involved in the application of high (in the biological context) pressure.

Intermolecular Interactions

The native structure contains 12 intermolecular hydrogen bonds all of which are

retained in the low humidity form. Three more such hydrogen bonds are present in the latter. Water bridges also help hold together protein molecules in the crystals. Each of the two forms contains 16 water molecules involved in intermolecular bridges. However, they are involved in 39 intermolecular water bridges in the low humidity form and 30 in the native form. Only 11 of the 16 water molecules in the two forms are common. These 11 water molecules are involved in 31 water bridges in the low humidity form and 20 in the native structure. Thus, not surprisingly, the number of intermolecular hydrogen bonds and water bridges is substantially larger in the low humidity form than in the native crystals.

Effect of the Transformation on the Hydration Shell

The low humidity form of tetragonal lysozyme provides an opportunity to examine the changes in the hydration shell of the protein caused by a small perturbation in the surrounding bulk water. The changes could be of two types: One, that in the number of water molecules that surround the protein molecule and two, that in the location of water molecules with respect to the protein molecule. Both types of changes were investigated by carefully comparing the water structure in the native and the low humidity tetragonal lysozyme. In both the structures a water molecule was considered as interacting with the protein molecule if it is at a distance of 3.6 \AA or less from a nitrogen or an oxygen atom in the protein molecule. Using this criterion, the number of water molecules that surround and interact with the protein is 121 in the low humidity form and 120 in the native crystals. Thus the number of water molecules in the first hydration shell is nearly the same in the two forms. In order to explore the second type of changes, a water molecule in one form and another in the other was considered as equivalent if they interacted with at least one common atom and if the distance between the two water molecules (after appropriate superposition) was less than 1.8 \AA . This criterion is somewhat arbitrary and rather liberal. When it is used, it turns out that 75 water molecules have equivalents. These may be considered, in a limited sense, as the invariant water molecules in the hydration shell. Thus more than a third of the water molecules have different locations in the two hydration shells. Even allowing for the inherent inaccuracies in water positions, this appears to be a significant observation.

The invariant water molecules in the low humidity form were superposed as a whole on the corresponding water molecules in the native form. The rotation and the translation necessary for doing so turns out to be nearly the same as those used for superposing the protein molecules in the two forms. The hydration shell as a whole thus appears to move along with the protein molecule during the water mediated transformation, despite the significant changes that occur in the location of many water molecules. Another general observation pertains to the number of protein–water interactions. This number increases from 205 in the native crystals to 223 in the low humidity form. It would thus appear that the hydration shell tends to become more strongly attached to the protein when the amount of water surrounding it decreases.

Predictably the water molecules in the first hydration shell most seriously affected during the water mediated transformation are those with only a single interaction with the protein. In the low humidity form, out of 64 such water molecules, 22 interact with main chain atoms and 42 with side chain atoms. A little less than half (30) of these singly bound water molecules have no equivalents. Among these, 7 are attached

Table 2. Distribution of water molecules and average B values in different ranges of the distance between equivalent water molecules.

Distance between equivalents	No of water molecules	Mean B
0.00–0.25 Å	15	21(9)Å ²
0.25–0.50	31	30(14)
0.50–1.00	19	33(10)
1.00–1.80	12	43(9)

to main chain atoms and 23 to side chain atoms. Thus, the proportion of water molecules without equivalents is higher among those bound to side chains than among those bound to the main chain. Of the 31 water molecules with two protein interactions in the low humidity form, only a third (10) are devoid of equivalents in the native crystals. Interactions in eight of these 10 exclusively involve side chain atoms. Only two of the 26 water molecules with three or more interactions with the protein molecule in the low humidity form are devoid of equivalents in the native crystals. One of them has three interactions with side chains while the other has one interaction with a side chain and two interactions with the main chain. Both of them have high temperature factors.

The B value of a protein-bound water oxygen may, in a very loose sense, be considered to depend in part upon the strength of its interaction with the protein molecule. It is therefore interesting to examine the distribution of water molecules and average B values in the low humidity form in different ranges of the distance between equivalent water molecules. The distribution, given in table 2, peaks in the 0.25 to 0.50 Å region. In fact for half the relevant water molecules, the distance between equivalents is less than 0.5 Å. As can be seen from the table, on an average the temperature factor steadily increases with the distance between equivalents. The average B value of protein-bound water molecules with no equivalent is still higher at 50 Å².

Equivalent water molecules in the two structures do not always have identical sets of interactions. For instance, among the 21 doubly bound water molecules which have equivalents, both the interactions with the protein are the same in the two crystal forms for 14 while only one interaction is the same for the remaining 7. The 24 multiply bound (three or more interactions) water molecules in the low humidity form, which have equivalents in the native crystals, are involved in a total of 96 interactions. Of these 81 are retained in the native crystals.

Conclusion and Prospects

Water-mediated transformation caused by the gentle removal (or addition) of a small amount of bulk water from the protein crystals when the relative humidity of the environment is systematically varied, is a widespread though not universal phenomenon. It appears to provide a novel tool for exploring the variability in protein hydration and its structural consequences. The X-ray analysis of the low humidity structure of tetragonal lysozyme indicates that a change in the quantity of bulk water

produces significant changes in the first hydration shell which, as a whole, tends to move along with the protein molecule. Understandably, the water molecules least affected by the changes are those with strong multiple interactions with the protein. The effect of the disturbance in the bulk water increases as the number of interactions with the protein molecule and the strength of interactions decrease. The effect is higher when the interactions involve side chains, especially in the case of singly or doubly bound water molecules. The changes in the hydration shell cause structural perturbations in the molecule. In the case of lysozyme, these perturbations are most pronounced in regions which move during substrate binding.

The structure analysis of low humidity tetragonal lysozyme represents the first step in a long range programme of X-ray studies on the low humidity forms of a few well known protein crystals. The most spectacular water-mediated transformation in protein crystals observed so far is in monoclinic lysozyme. The native crystals contain two crystallographically independent molecules related to each other by a pseudo B centring (27). The transformation takes place between relative humidities of 93 and 90%. The two molecules become equivalent in the low humidity form thus making the structure truly B centred (22). This cell which has a substantially lower volume on account of the loss of solvent can be transformed as illustrated in figure 22, into a primitive cell with identical a^* and c^* directions. The diffraction patterns from the native and low humidity forms are given in figure 23. The solvent content of the low humidity monoclinic lysozyme at 22% by volume is perhaps the lowest observed so far among protein crystals. The crystals diffract much better at low humidity than in their native state. The structure of the low humidity form has been determined recently and refined to an R value of 0.175 for 7660 reflections in the 10–1.75 Å resolution shell (61). The refined structure is currently being carefully examined. Also under investigation are the low humidity forms of ribonuclease A (21, 22). These studies based on the novel approach involving water-mediated transformation, taken

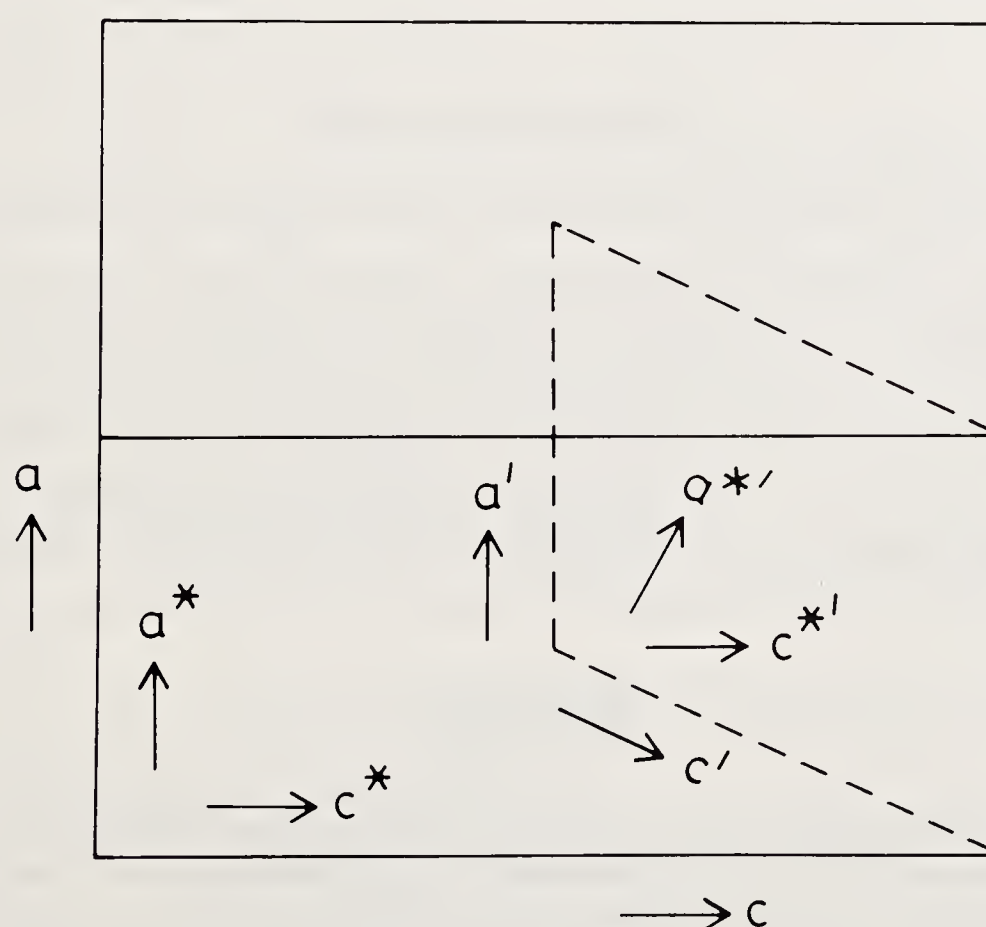


Figure 22. The relationship between the unit cells of the high and low humidity forms of monoclinic lysozyme. The primed symbols correspond to the low humidity form. Taken from ref. 22.

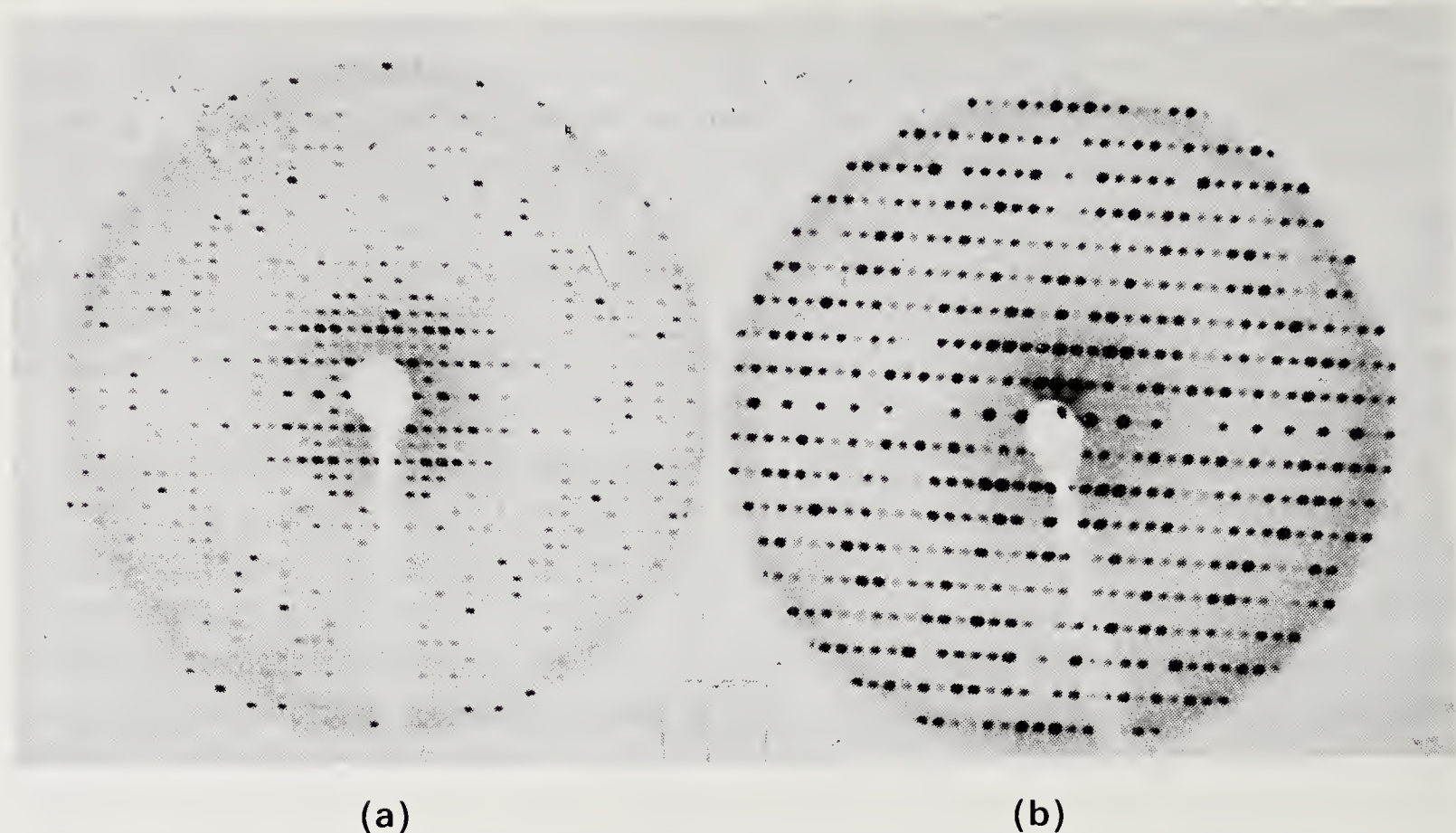


Figure 23. 15 precession photographs from monoclinic lysozyme at (a) 100% and (b) 90% r.h. Taken from ref. 22.

together, are expected to provide a better understanding of protein hydration, the nature and the extent of the possible variation in it, and the influence of this variation on the structure and action of proteins. The approach involving water-mediated transformations has been shown to be useful also in delineating the rigid and the flexible regions in protein molecules and in identifying the invariant features in their hydration shells (62).

Acknowledgements

We thank the Department of Science & Technology, Government of India, for financial support. The work reviewed and discussed here has had contributions from D. M. Salunke, B. Veerapandian and C. G. Suresh in the past. We thank Sir David Phillips for encouragement and for providing us with atomic parameters and structure factors of native tetragonal lysozyme. In the course of the work Eleanor Dodson, M. R. N. Murthy and C. Ramakrishnan have helped us with computer programmes. We have received technical help from C. Govindswamy and K. B. Shobana.

References

1. Kunz Jr., I. D. & Kauzmann, W. (1974) *Adv. Protein Chem.* **28**, 239–345.
2. Careri, G., Gratton, E., Yang, P. H. & Rupley, J. A. (1980) *Nature (London)* **285**, 572–573.
3. Poole, P. L. & Finney, J. L. (1983) *Biopolymers* **22**, 255–260.
4. Baker, L. J., Hansen, A. M. F., Bhaskara Rao, P. & Bryan, W. L. (1983) *Biopolymers* **22**, 1637–1640.
5. Finney, J. L., Goodfellow, J. M., Howell, P. L. & Vovelle, F. (1985) *J. Biomol. Struct. Dynam.* **3**, 599–622.

6. Piculles, L. & Hallel, B. (1986) *J. Chem. Soc., Faraday Trans. 1*, **82** (2), 401–414.
7. Finney, J. L. (1979) in *Water, A Comprehensive Treatise*, ed. Franks, F. (Plenum Press, New York), Vol. 6, pp. 47–122.
8. Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1979) *J. Mol. Biol.* **131**, 509–522.
9. Sakabe, K., Sakabe, N. & Sasaki, K. (1980) in *Water and Metal Cations in Biological Systems*, eds. Pullman, B. & Yagi, K. (Japan Scientific Societies Press, Tokyo), pp. 117–128.
10. Blake, C. C. F., Pulford, W. C. A. & Artymiuk, P. J. (1983) *J. Mol. Biol.* **167**, 693–723.
11. Blundell, T. L., Barlow, D. J., Borkakoti, N. & Thornton, J. M. (1983) *Nature (London)* **306**, 281–283.
12. Baker, E. N. & Hubbard, R. E. (1984) *Prog. Biophys. Mol. Biol.* **44**, 97–179.
13. Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N. & Vijayan, M. (1988) *Phil. Trans. R. Soc. (London)* **319**, 369–456.
14. Thanki, N., Thornton, J. M. & Goodfellow, J. M. (1988) *J. Mol. Biol.* **202**, 637–657.
15. Baker, E. N. (1988) *J. Mol. Biol.* **203**, 1071–1095.
16. Crowfoot, D., Riley, D. Bernal, J. D., Fankuchen, I. & Perutz, M. (1938) *Nature (London)* **141**, 521–524.
17. Crowfoot, D. & Riley, D. (1939) *Nature (London)* **144**, 1011–1012.
18. Watson, B. J., Davidson, E. & Perutz, M. F. (1947) *Proc. R. Soc. London A* **191**, 83–132.
19. Huxley, H. E. & Kendrew, J. C. (1953) *Acta Crystallogr.* **6**, 76–80.
20. Perutz, M. F. (1954) *Proc. R. Soc. A* **225**, 264–286.
21. Salunke, D. M., Veerapandian, B. & Vijayan, M. (1984) *Curr. Sci.* **53**, 231–235.
22. Salunke, D. M., Veerapandian, B., Kodandapani, R. & Vijayan, M. (1985) *Acta Crystallogr.* **B41**, 431–436.
23. Kartha, G., Bello, J. & Harker, D. (1967) *Nature (London)* **213**, 862–865.
24. Carlise, C. H., Palmer, R. A., Muzumder, S. K., Gorinsky, B. A. & Yeates, D. G. R. (1974) *J. Mol. Biol.* **85**, 1–18.
25. Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1965) *Nature (London)* **206**, 757–761.
26. Artymiuk, P. J., Blake, C. C. F., Rice, D. W. & Wilson, K. S. (1982) *Acta Crystallogr.* **A38**, 778–783.
27. Hogle, J., Rao, S. T., Mallikarjunan, M., Beddell, C., McMullar, R. K. & Sundaralingam, M. (1981) *Acta Crystallogr.* **B37**, 591–597.
28. Moulton, J., Yonath, A., Traub, W., Smilansky, A., Podjarny, A., Rabinovich, D. & Sayer, A. (1976) *J. Mol. Biol.* **100**, 179–195.
29. Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimer, B. & Sheat, S. (1969) *Nature (London)* **224**, 491–495.
30. Bentley, E., Dodson, E., Dodson, G., Hodgkin, D. & Mercola, D. (1976) *Nature (London)*, **261**, 166–168.
31. Hardman, K. D., Wood, M. K., Schiffer, M., Edmundson, A. B. & Ainsworth, C. F. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 1393–1397.
32. Rockland, L. B. (1960) *Anal. Chem.* **32**, 1375–1376.
33. Kodandapani, R., Suresh, C. G. & Vijayan M. (1990) *J. Biol. Chem.* **265**, 16126–16131.
34. Phillips D. C. personal communication.
35. Konnert, J. H. (1976) *Acta Crystallogr.* **A32**, 614–617.
36. Hendrickson, W. A. & Konnert, J. H. (1980a) in *Computing in Crystallography*, eds. Diamond, R., Ramaseshan, S. & Venkatesan, K. (Indian Academy of Sciences, Bangalore), pp. 13.01–13.26.
37. Hendrickson, W. A. & Konnert, J. H. (1980b) in *Biomolecular Structure, Function, Conformation and Evolution*, ed. Srinivasan, R. (Pergamon Press, Oxford) Vol. 1, pp. 43–57.
38. Luzzati, V. (1952) *Acta Crystallogr.* **5**, 802–810.
39. IUPAC-IUB Commission on Biochemical Nomenclature (1970) *J. Mol. Biol.* **52**, 1–17.
40. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1968) *J. Mol. Biol.* **7**, 95–99.
41. Janin, J., Wodak, S., Levitt, M. & Maigret, B. (1978) *J. Mol. Biol.* **125**, 357–386.
42. Bhat, T. N., Sasisekharan, V. & Vijayan, M. (1979) *Int. J. Pept. Protein Res.* **13**, 170–184.
43. McGregor, M. J., Islam, S. A. & Sternberg, M. J. E. (1987) *J. Mol. Biol.* **198**, 295–310.
44. Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C. & Rupley, J. A. (1972) in *The Enzymes*, ed. Boyer, P. D. (Academic Press, New York, London), **7**, pp. 665–808.

45. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **3**, 284–438.
46. Bansal, M. (1976) Ph. D. Thesis, Indian Institute of Science, Bangalore, pp. 220–224.
47. Schulz, G. E. & Schirmer, R. H. (1979) *Principles of Protein Structure*, (Springer-Verlag, New York Inc., Amsterdam, North Holland, Academic Press), pp. 69.
48. Richardson, J. S., Getzoff, E. D. & Richardson, D. C. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2574–2578.
49. Richardson, J. S., (1981) *Adv. Protein Chem.* **34**, 167–339.
50. Nair, C. M. & Vijayan, M. (1980) *J. Chem. Soc., Perkin Trans. 2*, 1800–1804.
51. Matthews, B. W. (1972) *Macromolecules* **5**, 818–819.
52. James, E., Milner, W., Roslan, I., Khaled, B. M. & Ron, P. (1988) *J. Mol. Biol.* **204**, 777–782.
53. Salunke, D. M. & Vijayan, M. (1981) *Int. J. Pept. Protein Res.* **18**, 348–351.
54. Rossmann, M. G. & Argos, P. (1975) *J. Biol. Chem.* **250**, 7525–7532.
55. Kundrot, C. E. & Richards, F. M. (1987) *J. Mol. Biol.* **193**, 157–170.
56. Johnson, L. N. & Phillips, D. C. (1965) *Nature (London)* **206**, 157–170.
57. Ford, L. O., Johnson, L. N., Machin, P. A. & Phillips, D. C. (1974) *J. Mol. Biol.* **88**, 349–371.
58. Perkins, S. J., Johnson, L. N., Machin, P. A. & Phillips, D. C. (1978) *Biochem. J.* **A173**, 607–616.
59. Kelly, J. A., Sielecki, A. R., Sykes, B. D., James, M. N. G. & Phillips, D. C. (1979) *Nature (London)* **282**, 875–878.
60. Kundrot, C. E. & Richards, F. M. (1988) *J. Mol. Biol.* **200**, 401–410.
61. Madhusudan, Kodandapani, R. & Vijayan, M. (unpublished results).
62. Madhusudan & Vijayan, M. (1991) *Curr. Sci.* **60**, 165–170.

Structure of Bowman–Birk Protease Inhibitor and Its Complex with Trypsin

Atsuo Suzuki, Yasuo Tsunogae, Takashi Yamane and
Tamaichi Ashida¹

Department of Applied Chemistry, School of Engineering, Nagoya University,
Nagoya, 464-01 Japan

Introduction

Bowman–Birk type protease inhibitors isolated from the leguminous seeds are small proteins made of 60 to 80 amino acid residues, which are active against serine proteases. They have several unique features:

- (i) Presence of several different but homologous inhibitors in each species,
- (ii) Unusually large content of seven disulfide bridges at genetically conserved sites,
- (iii) Contains very few hydrophobic residues other than proline,
- (iv) Two domains of structure made by tandem repeat of homologous amino acid sequences within a single chain,
- (v) Double-headed inhibitors, each domain having a reactive site against proteases, typically trypsin and chymotrypsin (1).

It is known that the inhibitors are accumulated in a short period of the final stage of the development of the dicot seed on the mother plant (2), and upon germination and during early growth of the seedling these inhibitors are consumed to supply energy and biosynthetic intermediates to the growing plant (3). The inhibitors are appropriate as storage proteins since they provide resistance against the proteases of insects and microorganisms. They may inhibit expression of the activities of unknown proteases in the seeds during their dormancy which initiate germination. But their precise biological functions are still not clear. The molecular evolution of this type of inhibitors has been discussed (4). The crystal structure study of these inhibitors and their complexes with proteases is very important from the viewpoint of both structural chemistry and biology.

We have obtained crystals of several inhibitors and their complexes with proteases (5, 6). Large crystals of AB-I from adzuki bean were obtained but their crystallinity was too poor to give X-ray diffraction data good enough for high resolution structure analysis; however, the (1:1) complex of AB-I with bovine trypsin gave much better crystals (7, 8). The A-II inhibitor from peanut also gave a promising crystal (9). The

¹To whom all correspondence should be addressed.

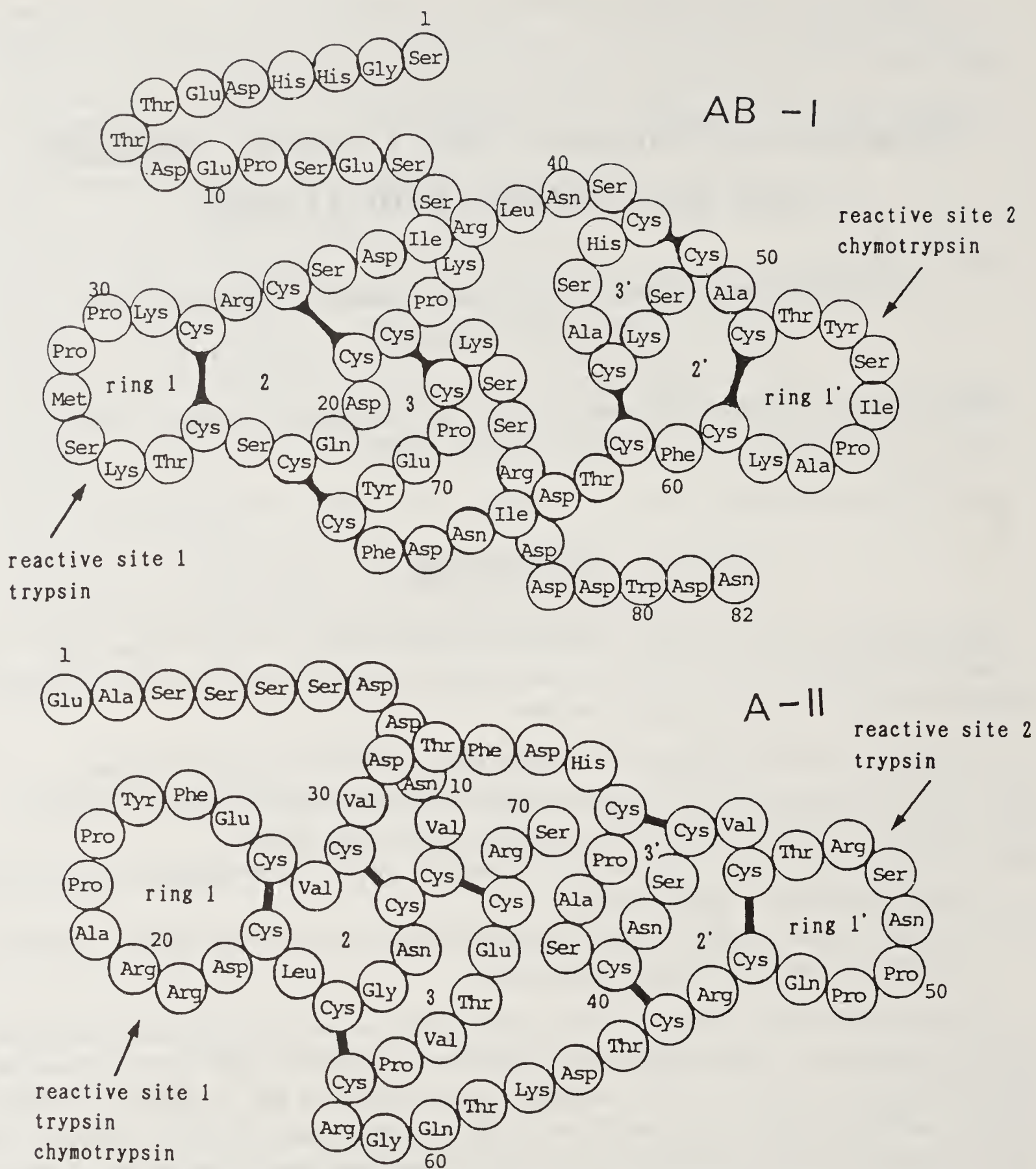


Figure 1. The primary structure of AB-I (10) and A-II (4). The reactive sites and the ring numbering are shown.

structure analyses of both these crystals have successfully been carried out at 2.3 Å resolution.

The primary amino acid sequences of AB-I and A-II are shown in figure 1. AB-I consists of 82 amino acid residues with molecular weight 9,100, the active sites are Lys 26 for trypsin and Tyr 53 for chymotrypsin (10). A-II consists of 70 amino acid residues with molecular weight 7,600, Arg 19 inhibits both trypsin and chymotrypsin, while Arg 47 inhibits trypsin (4, 11). The peanut inhibitors have a unique amino acid sequence among this family of the inhibitors:

- (i) Two insertions in its reactive ring 1 of domain 1,
- (ii) Replacement of P1' site, usually serine with arginine,
- (iii) The presence of the Pro-Pro sequence in ring 1' of domain 2.

Materials and Methods

Trypsin·AB-I complex crystal: The AB-I inhibitor was extracted from adzuki bean ‘Takara’ (*Vigna angularis*) and purified according to the reported procedure (10). Purified AB-I and bovine trypsin were mixed at (1:1:1·0) molar ratio and purified by gel chromatography; the molecular weight of the complex was close to the sum of the molecular weights of AB-I and trypsin (32,400). The hanging drop vapor diffusion method was successfully applied for crystallization. Crystals were tetragonal bipyramids with maximum dimensions of $0·5 \times 0·4 \times 0·4 \text{ mm}^3$: space group $P4_12_12$, $a = 55·42$, $c = 181·72 \text{ Å}$, one (1:1) complex per asymmetric unit. The solvent content was 39% (w/w). The inhibitory activity assay and the amino acid analysis of the recovered materials from the crystals showed that they really consist of trypsin and AB-I, but the N-terminal 8 residues (Ser 1–Thr 8) and C-terminal 6 residues (Asp 77–Asn 82) are lost; both are known to be susceptible to a protease attack (8). Only a heavy atom replaced crystal of $[\text{UO}_2(\text{NO}_3)_2]$ was found to be available for the phase determination.

A-II crystal: A-II was isolated from peanut (*Arachis hypogaea*) and purified as described previously (11). The crystals were hexagonal bipyramids with maximum dimensions of $1·5 \times 1·0 \times 1·0 \text{ mm}^3$: $P3_121$, $a = 71·8$, $c = 65·9 \text{ Å}$. The asymmetric unit contains two inhibitor molecules with the solvent content 61% (w/w). Two isomorphous heavy atom crystals of $\text{Hg}(\text{AcO})_2$ and K_3IrCl_6 were obtained.

X-ray intensity measurement: X-ray intensity data for high resolution study were collected with X-rays from synchrotron radiation at Photon Factory, National Laboratory for High Energy Physics. Arndt & Wonacott oscillation camera and Sakabe’s Weissenberg camera (12) were used, the intensity data for A-II were recorded on IP (Fuji imaging plate). For both crystals the intensity data with $d < 2·3 \text{ Å}$ are too weak and were not available for the structure analysis.

Structure determination, trypsin·AB-I complex: A 3 Å resolution electron density map was synthesized, making use of the single isomorphous (SIR) and anomalous data of the U-derivative. The map was poor and the peptide chain could not be traced. But the translation search for the oriented trypsin in the electron density map easily gave a solution, the orientation was obtained from the Patterson map. The initial SIR phases were combined with those calculated from the trypsin model, and a considerably improved map was obtained, continuous high density regions appeared at the site where the inhibitor molecule was expected to be located. It is to be noted that although the contribution of trypsin to the structure factor is dominative, the Fourier map based on the trypsin phases could not show up AB-I. The map was further improved by an iterative Fourier and phase combining procedure; the solvent region was flattened and the map was Fourier-inverted to give new phases (13), which were again combined with the original SIR phases. After seven cycles of this procedure, an atomic model of the trypsin-binding domain of AB-I could be built with the program PMAKER (14). The electron density corresponding to the rest of AB-I could not be found, suggesting that it may be in a highly disordered state. A part of the final electron density map showing a *cis* peptide bond of Met 28–Pro 29 is shown in figure 2. The structure was refined by the Hendrickson–Konnert least-squares

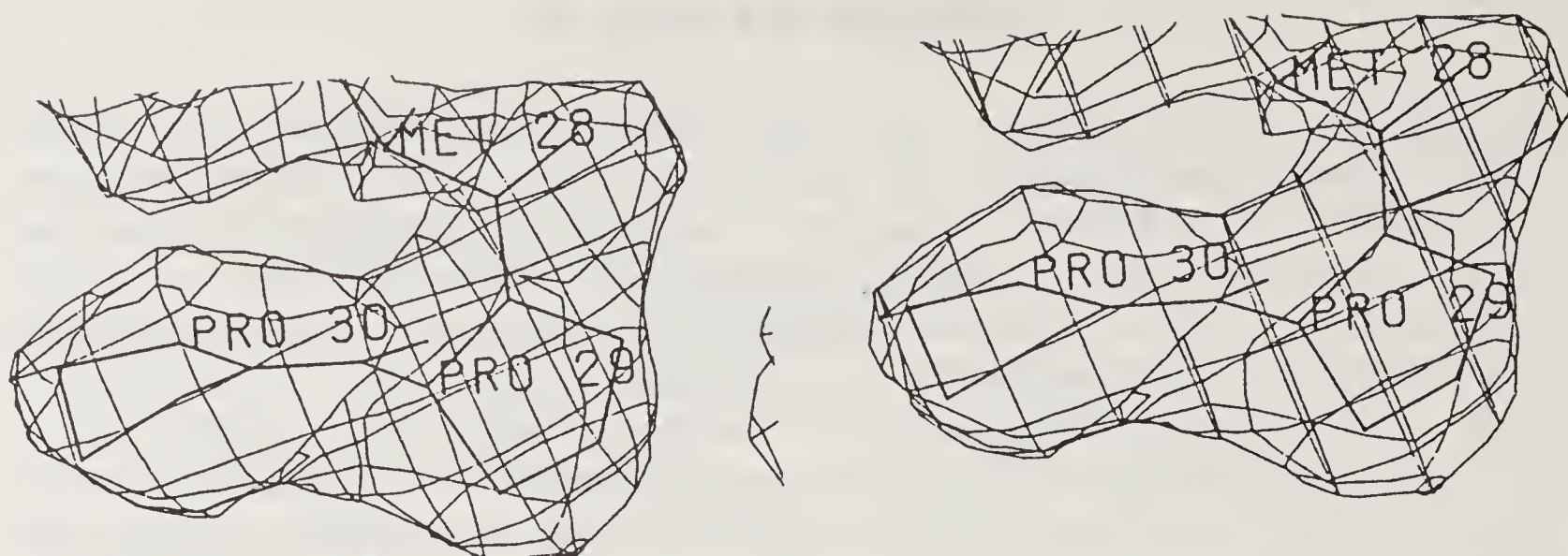


Figure 2. A part of the electron density map of the trypsin·AB-I complex crystal, showing a *cis* peptide bond, Met 28–Pro 29 in AB-I (stereo).

procedure (15); the atoms included in the refinement are: 1,629 atoms of 223 residues of trypsin, 275 atoms of 36 residues of AB-I (Ser 12–Arg 38 and Asn 65–Lys 73) and 140 water molecules. Number of reflections is 7,188 [$6 > d > 2.3 \text{ \AA}$, $F_0 > 3\sigma(F_0)$], number of variables is 8,177 and $R = 0.20$ (13).

Structure determination, A-II: The first 6 \AA resolution Fourier map with the MIR phases showed a good contrast between protein and solvent regions. Only the SIR phases with anomalous effect could be assigned for the reflections with $6 > d > 3.3 \text{ \AA}$, giving a rather noisy Fourier map. A-II exists in the crystal as a tetramer with a local D_2 symmetry. The averaged electron densities of two crystallographically independent molecules were Fourier-inverted to give new phases. The new phases were combined with the original ones, and a new improved map was synthesized, in which the polypeptide chain could easily be traced. By the averaging process the map was considerably improved. Then phasing was extended to $d = 2.8 \text{ \AA}$ by Fourier transform of the electron density map. The model was built with the aid of PMAKER, and refined by the Hendrickson–Konnert program with frequent interruption by model adjustment. For both independent molecules the electron densities for the N-terminal 9 residues and the C-terminal 2 residues are so weak that no model could be built, while the conformations of almost all the side chains including long chains such as Lys and Arg within the residues 10–68 could be determined. The refinement included 890 atoms of two A-II molecules and 54 solvents with 7,697 reflections [$6 > d > 2.3 \text{ \AA}$, $F_0 > 3\sigma(F_0)$], $R = 0.22$ (16).

Results and Discussion

Crystallinity: Among five different Bowman–Birk type inhibitors isolated from peanut, A-I, A-II, B-I and B-III have the same amino acid sequence except for differences in the N-terminal region. A-II consisting of 70 amino acid residues is the largest, A-I consists of Ser 4–Ser 70 of A-II, B-I Asp 8–Ser 70, and B-III is the smallest with Val 10–Ser 70 (4). The crystals of A-I, A-II and B-I are isomorphous with each other. This seems to be quite reasonable since the N-terminal peptide chain

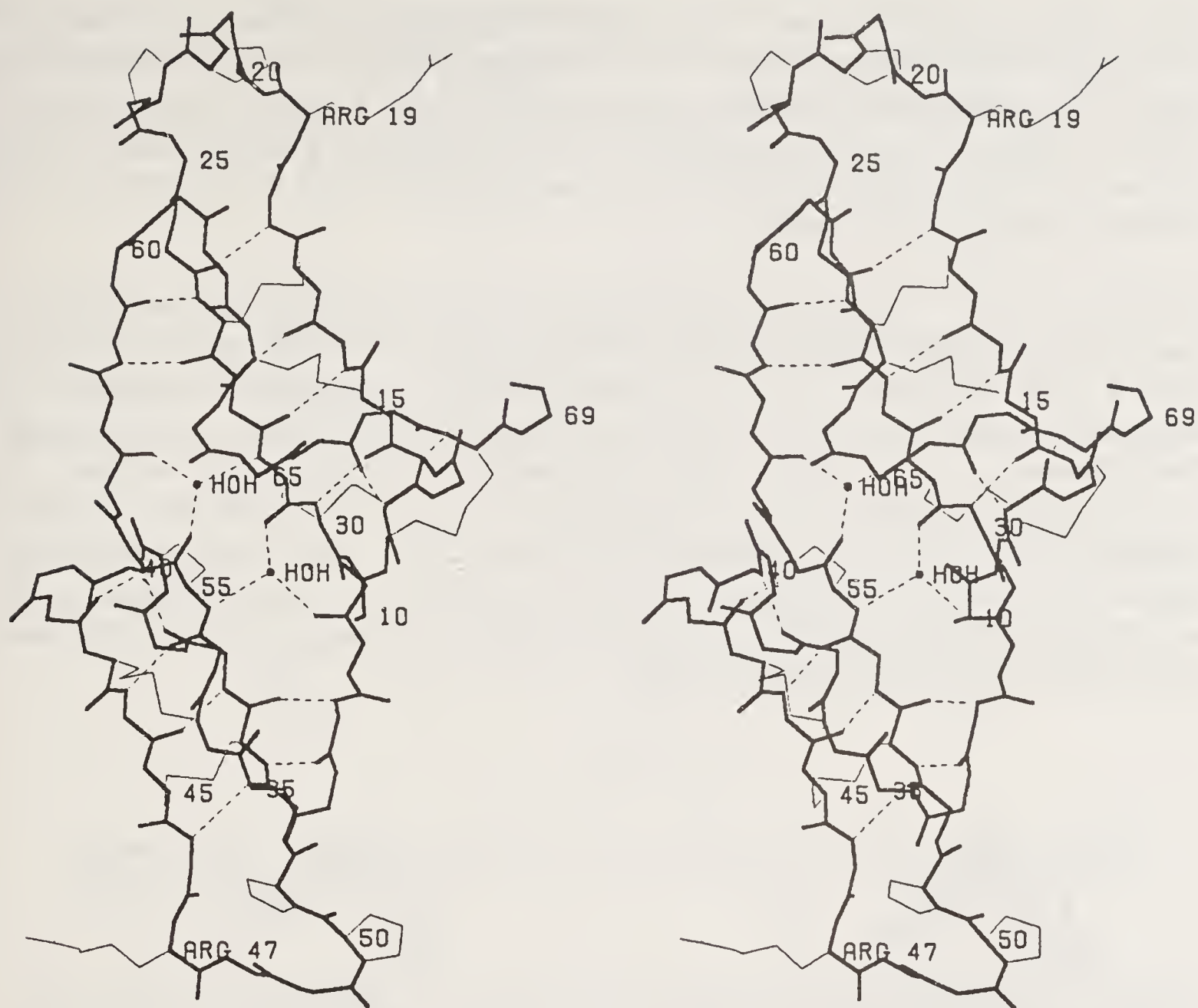


Figure 3. A stereo drawing of the peptide main chain of A-II, also showing the reactive Arg at sites 19 and 47, two Pro-Pro sequences in each reactive site, inter-main chain hydrogen bonds, and two water molecules within the β sheet each having three hydrogen bonds to the peptide chains.

of Glu 1-Asn 9 is highly hydrophilic, and will take many random conformations in the solvent region as was suggested by the electron density map of A-II, in which the electron density corresponding to the N-terminal chain is at the level of the solvent and could not be identified at all. B-III which does not have a long hydrophilic chain at any terminus gives crystals of the highest quality among this family. AB-I and AB-IIa from adzuki bean both have long hydrophilic peptide chains at both termini, their crystallinity being considerably worse than that of A-II. We had expected that the crystals of this family of the inhibitors would be of higher quality than many other proteins, since they have unusually many disulfide linkages in such small protein molecules. But this is not the case. Each domain is rigid, but the long hydrophilic peptide chains at both termini require unusually large rooms in the solvent regions in order to accommodate many random conformations. Thus the intermolecular forces per unit volume are smaller than usual, and the crystals are of low quality.

In the electron density map of trypsin·AB-I complex only the trypsin binding domain of AB-I and trypsin could be found. Although there is enough room for the chymotrypsin binding domain, the electron density corresponding to it could not be

found. AB-I has Arg 38 in one of two peptide segments connecting two domains. If the peptide linkage following Arg 38 was hydrolyzed during the binding reaction with trypsin, the chymotrypsin binding domain would be in a highly disordered state. Then the height of the electron density of the chymotrypsin binding domain in the map would not be higher than that of the solvent region, and determination of its structure would not be possible.

Structure of A-II: The peptide main chain with disulfide linkages and possible inter-main chain hydrogen bonds of A-II is shown in figure 3, and all the side chains in figure 4. The structures of two crystallographically independent molecules are essentially the same, though two molecules are independently refined at the final stage. The differences of the main chain conformational angles (ϕ, ψ) between the corresponding residues of the two independent molecules are small, the maximum difference is 29° , and the mean differences in ϕ and ψ are 8° and 9° , respectively. Some residues have different conformations of the side chains due to the differences in the surroundings; the side chains of almost all the residues including long hydrophilic ones can be seen in the electron density map.

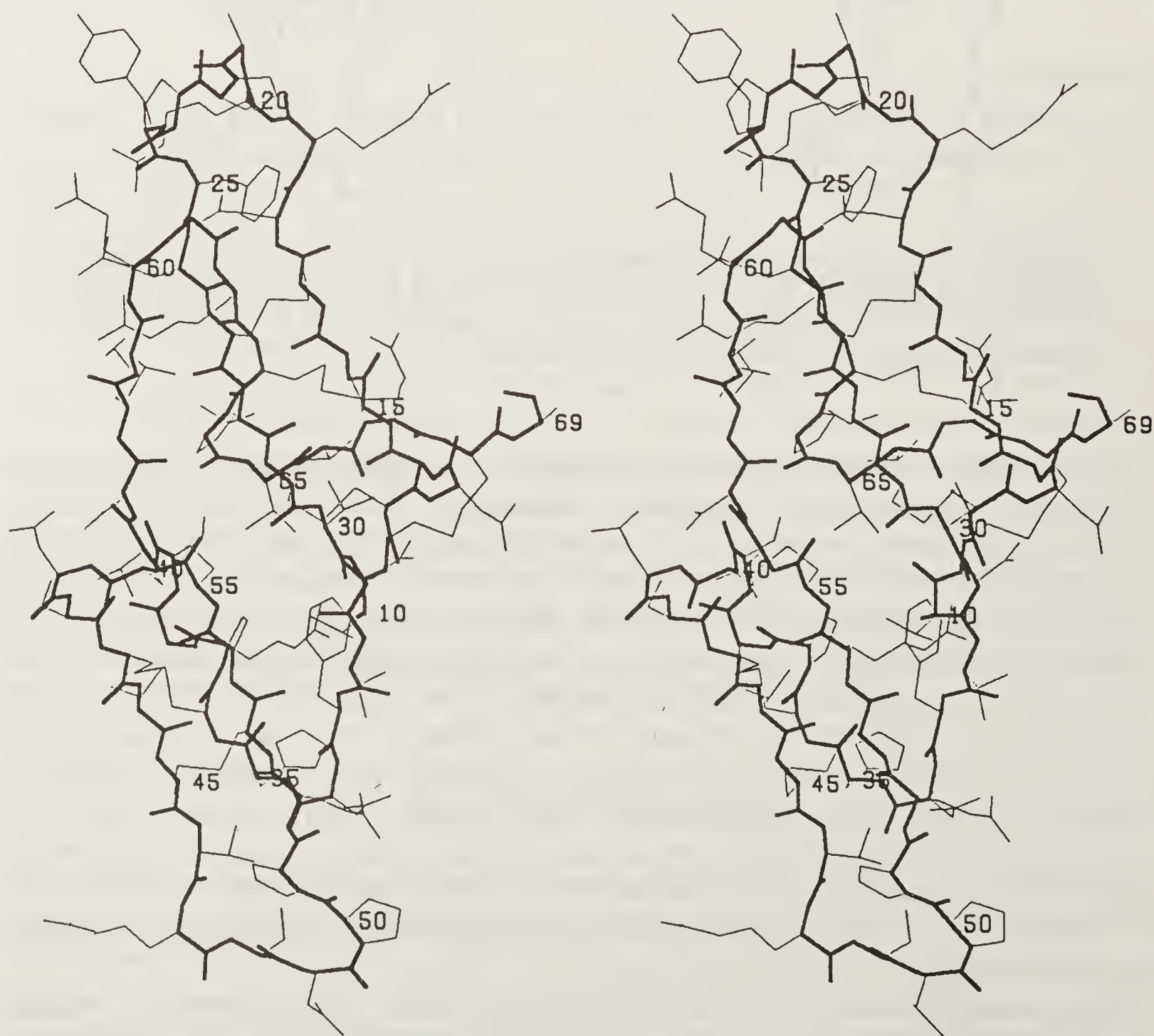


Figure 4. A stereo drawing of all the residues of A-II.

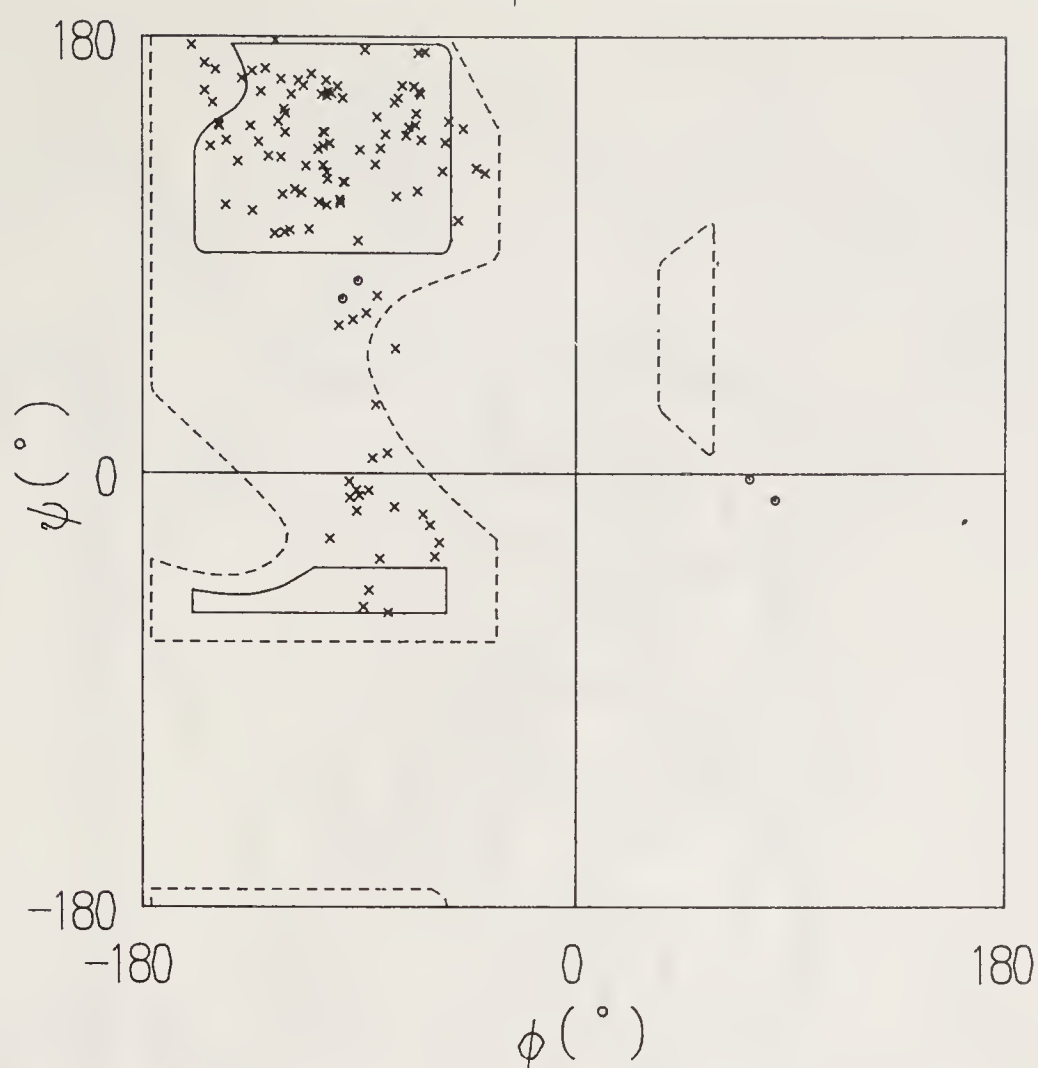


Figure 5. Ramachandran plot of A-II (two molecules). Those enclosed with small circles are Gly.

The molecule is in elongated form with dimensions of roughly $45 \times 20 \times 20 \text{ \AA}^3$, and consists of two distinct domains. The molecule itself has an intramolecular pseudo two-fold axis at the center as shown in figure 3; the structures of two domains are very similar to each other. Each domain consists of three rings made by disulfide linkages, rings 1 and 2 are roughly coplanar and make a β sheet at the bottom of the molecule, ring 3 is roughly perpendicular to the bottom. The outermost rings, 1 and 1', have reactive sites at both ends of the long molecule, thus the inhibitor is able to work as a divalent inhibitor. In figures 3 and 4 two long side chains at the upper right and lower left corners being active Arg 19 and Arg 47, these reactive sites are very convenient for proteases to recognize and to make a close contact to it.

The Ramachandran plot of two independent molecules is shown in figure 5. The conformational angles of ring 1 and 1' are listed in table 1. In this inhibitor most residues are in rather extended forms, with the conformational angles (ϕ, ψ) in the extended chain region. In the molecule neither α helices nor β turns exist. Two peptide bonds Pro 22–Pro 23 in domain 1 and Asn 49–Pro 50 in 2 are 'cis'. Four antiparallel peptide strands at the bottom of the molecule make a β sheet as shown in figure 3. Several hydrogen bonds around ring 2 make the main frame of domain 1, and those around ring 2' make the main frame of domain 2. In the vicinity of the center of the molecule the inner two peptide strands are separated too far from each other for hydrogen bond formation, but two water molecules are located midway between the two strands and each connects them via three hydrogen bonds. This is a unique β sheet structure not reported so far. This four-strand β sheet still shows a common left-handed twist when viewed along the hydrogen bond direction. Thus the two domains are connected not only by two peptide segments, but also by several hydrogen bonds

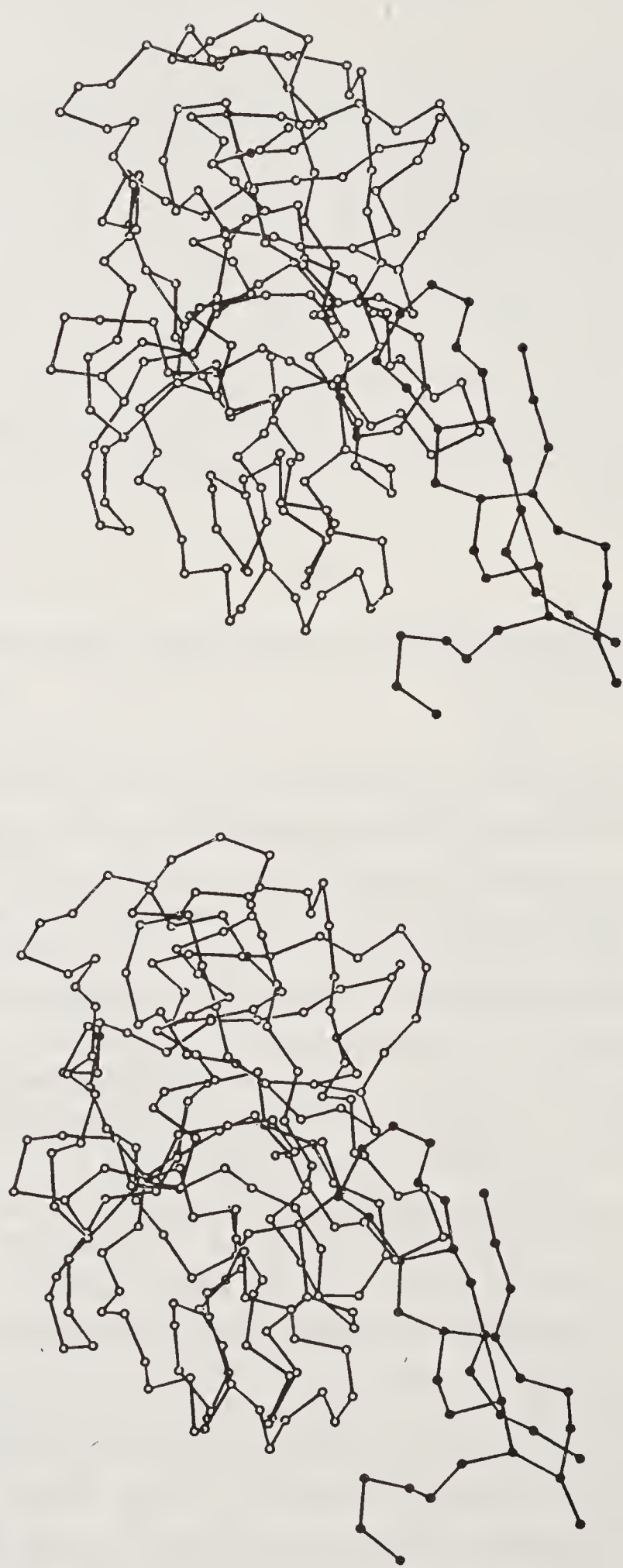


Figure 6. A stereo drawing of the C α atom model of the trypsin·AB-I complex, only the trypsin binding domain (filled circles) is seen for AB-I, open circles for trypsin.

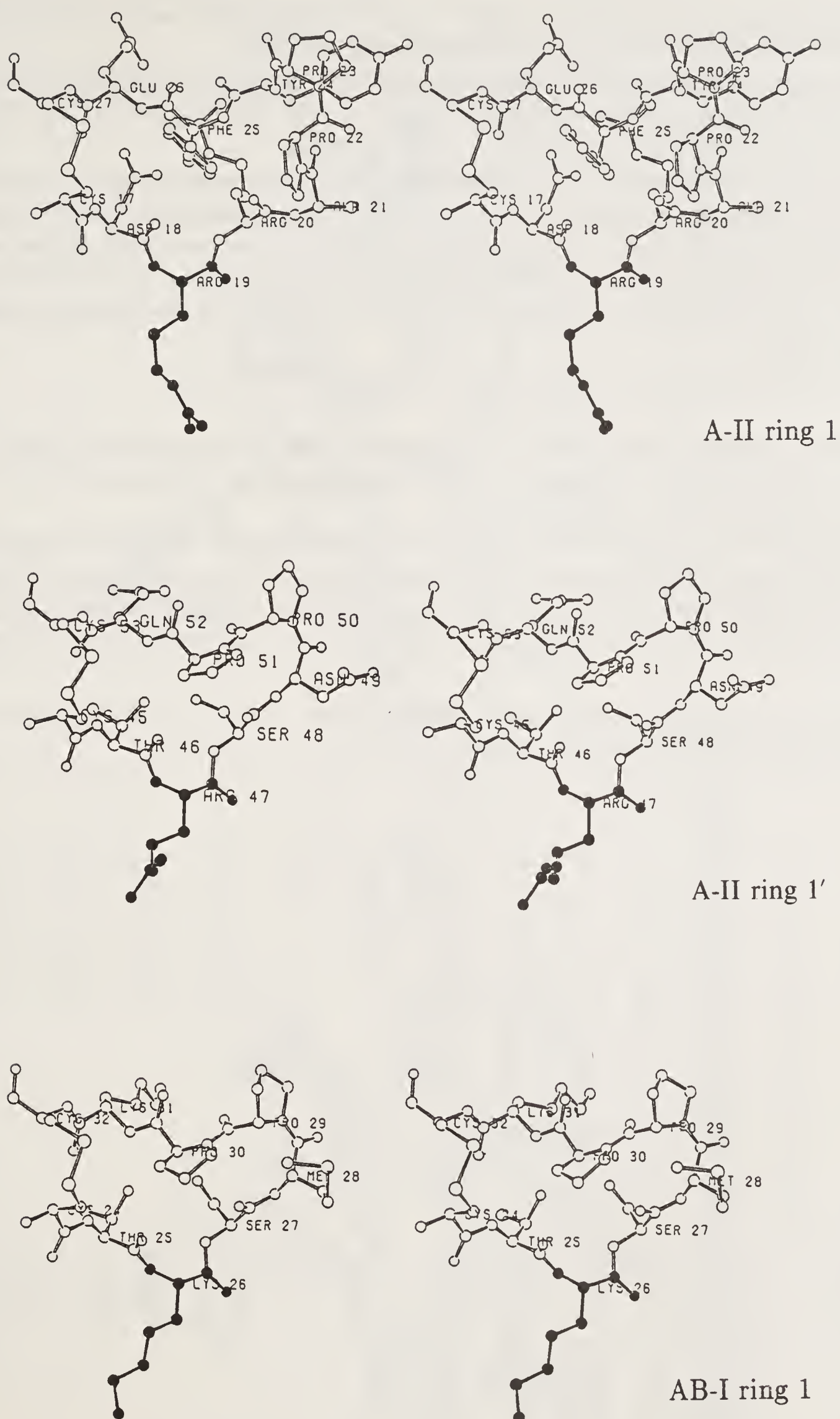


Figure 7. A comparison of the structures of ring 1: from top to bottom A-II ring 1, A-II ring 1', AB-I ring 1.

between the inner two peptide strands. A small hydrophobic core at the top of the molecule also plays a role in connecting the two domains.

The amino acid sequences suggest that from the viewpoint of genetics Tyr 24 and Phe 25 are undoubtedly the insertion in ring 1 of A-II (17), but the tertiary structure suggests that Pro 22–Pro 23 seem to be the insertion.

A-II exists in the crystal as a tetramer with an approximate symmetry D_2 , and the tetramer is made of two dimers, which are made with the interface formed as a parallel stack of two β sheets, bottom to bottom stacking of two monomers, with the main chain direction inclined at about 90° . Several hydrophobic side chains are between the two sheets. The main feature of the stacking is similar to that of *Streptomyces* subtilisin inhibitor (18).

The detailed structure of A-II will be reported elsewhere (16).

Structure of trypsin·AB-I complex: The reactive site of AB-I makes a close contact with the enzymatic active center of trypsin, giving rise to a (1:1) complex. The C^α model of the complex is shown in figure 6.

The structure of trypsin is not significantly different from that of free trypsin (19), some differences being observed in the enzymatic active center binding the inhibitor and in the regions which contact neighboring protein molecules in the crystal. The average deviation between the main chain atoms of the two tryptins after fitting by a

Table 1. Conformational angles (degrees) of ring 1 and 1' of A-II and AB-I and those of the reactive sites of inhibitors in their protease complexes.

		P3	P2	P1	P1'	P2'							(Ref.)
A-II 1	ϕ	−124	−74	−83	−132	−49	−66	−75	−103	−67	−105	−115	
	ψ	145	155	74	137	105	174	52	−27	144	141	162	
A-II 1'	ϕ	−113	−102	−93	−125	−146			−76	−65	−121	−99	
	ψ	160	157	64	179	111			153	158	100	160	
AB-I 1	ϕ	−99	−101	−112	−107	−122			−83	−66	−130	−119	
	ψ	176	168	46	171	106			151	178	123	139	
BPTI	ϕ	−77	−70	−117	−88	−113							(21)
	ψ	−29	156	39	164	79							
STI	ϕ	−46	−77	−89	—	−175							(22)
	ψ	−21	136	85	—	−179							
SSI	ϕ	−123	−52	−92	−117	−120							(20)
	ψ	141	134	89	169	78							
PSTI	ϕ	−118	−61	−94	−90	−96							(23)
	ψ	149	147	44	144	97							
A-II 1		Cys	Asp	Arg	Arg	Ala	Pro	cPro*	Tyr	Phe	Glu	Cys	
A-II 1'		Cys	Thr	Arg	Ser	Asn			cPro	Pro	Gln	Cys	
AB-I 1		Cys	Thr	Lys	Ser	Met			cPro	Pro	Lys	Cys	
BPTI		Pro	Cys	Lys	Ala	Arg							
STI		Ser	Tyr	Arg	Ile	Arg							
SSI		Cys	Pro	Met	Val	Tyr							
PSTI		Cys	Pro	Lys	Ile	Tyr							

The peptide bond between P1-P1' is the scissible bond.
A-II: free molecule, the values of one of two independent molecules are listed.
STI: the scissible bond is broken.
*cPro: the preceding peptide bond is *cis*.

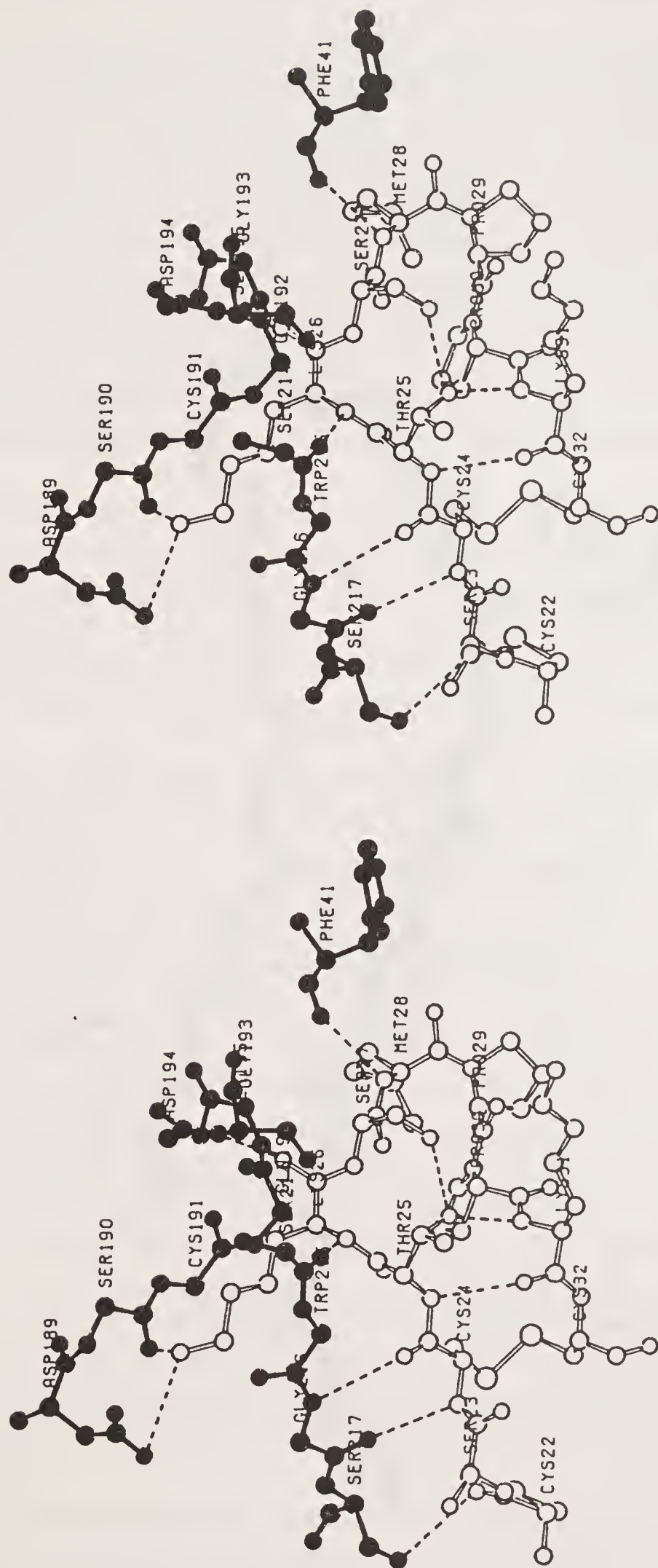


Figure 8. A stereo view of the contact region between the trypsin active center (filled circles) and the AB-I trypsin binding domain (open circles). Hydrogen bonds are shown with broken lines.

least-squares calculation is 0.69 Å. The temperature factors of the residues in contact with AB-I become smaller than those in the free enzyme. In fact the electron densities of the contact regions of both trypsin and the inhibitor are the highest and clearest in the map.

As mentioned before only the trypsin binding domain appeared for AB-I, its structure being very similar to those of A-II. The structures of the three rings, 1 and 1' of A-II and 1 of AB-I are compared in figure 7, and the conformational angles (ϕ, ψ) are listed in table 1 together with those of A-II. Especially the structure of ring 1 of present AB-I is essentially the same as that of ring 1' of A-II, both consist of 9 residues and have *cis* peptide bonds at the same sites, Met 28–Pro 29 in AB-I. The similarity between the two structures is striking, the biggest difference in the conformational angles is 23°. This structure may be the common one in the Bowman–Birk inhibitor family. Although ring 1 of A-II consists of 11 residues, the structure of its reactive site and its vicinity is also essentially the same as that of AB-I. This shows that the structure of the reactive site found in this study is very stable, and will not significantly be changed even on binding to proteases.

It is to be noted that the structure of the significant Pro–Pro sequences in AB-I ring 1 is the same as that of A-II ring 1', but is clearly different from that of A-II ring 1. It is not usual that such a large hydrophobic residue occupies the edge of the molecule which protrudes into solvents, but it certainly plays a part in keeping the structure of the ring in the stable form, and also a significant biological role, such as recognition of or by the target enzymes, its location being very appropriate for the latter role.

Trypsin·AB-I interactions: The structure of the contact region between trypsin and AB-I is shown in figure 8, and its schematic drawing in figure 9. The structures of the trypsin active site with and without Lys 26 of AB-I are compared in figure 10. There are extensive van der Waals contacts and many, about 15, hydrogen bonds between trypsin and AB-I. The side chain of Lys 26 elongates in the trypsin active pocket interacting with Asp 189, and the peptide chains of Cys 24–Lys 26 of AB-I and Ser 214–Gly 216 of trypsin make a small intermolecular antiparallel β sheet. Thus the contact between trypsin and AB-I is very close in a wide area. It is worthwhile to mention that the temperature factors of the atoms in the vicinity of the reactive site of the inhibitor decrease considerably on binding with the trypsin active center; figure 11 shows clearly such a tendency, while this is not found in A-II. This fact shows that the structure of the reactive site and its vicinity in the inhibitor becomes considerably more stable by forming the complex than that of the free molecule.

The Ser 195, O γ of trypsin comes in close contact with the C=O group of the scissible peptide bond of Lys 26–Ser 27 of AB-I; in the electron density map the two groups are continuous with high density, and the distance between Ser 195, O γ and Lys 26, C' is 2.3 Å, an intermediate between the covalent bond length and the van der Waals contact distance. The planarity of the C=O group, however, is yet preserved and the peptide bond is not broken.

In the case of a normal substrate, a conformational change from this complex state has to occur for the enzymatic reaction to proceed further. But the present structure is so stable that the occurrence of such a conformational change seems to be suppressed. The stability of the complex results not only from the contact between trypsin and AB-I but also from the structure of AB-I itself. The ring 1 has three intra-ring hydrogen bonds, Lys 31, N α –Thr 25, O γ , Ser 27, O γ –Thr 25, O γ and the β sheet hydrogen bond of

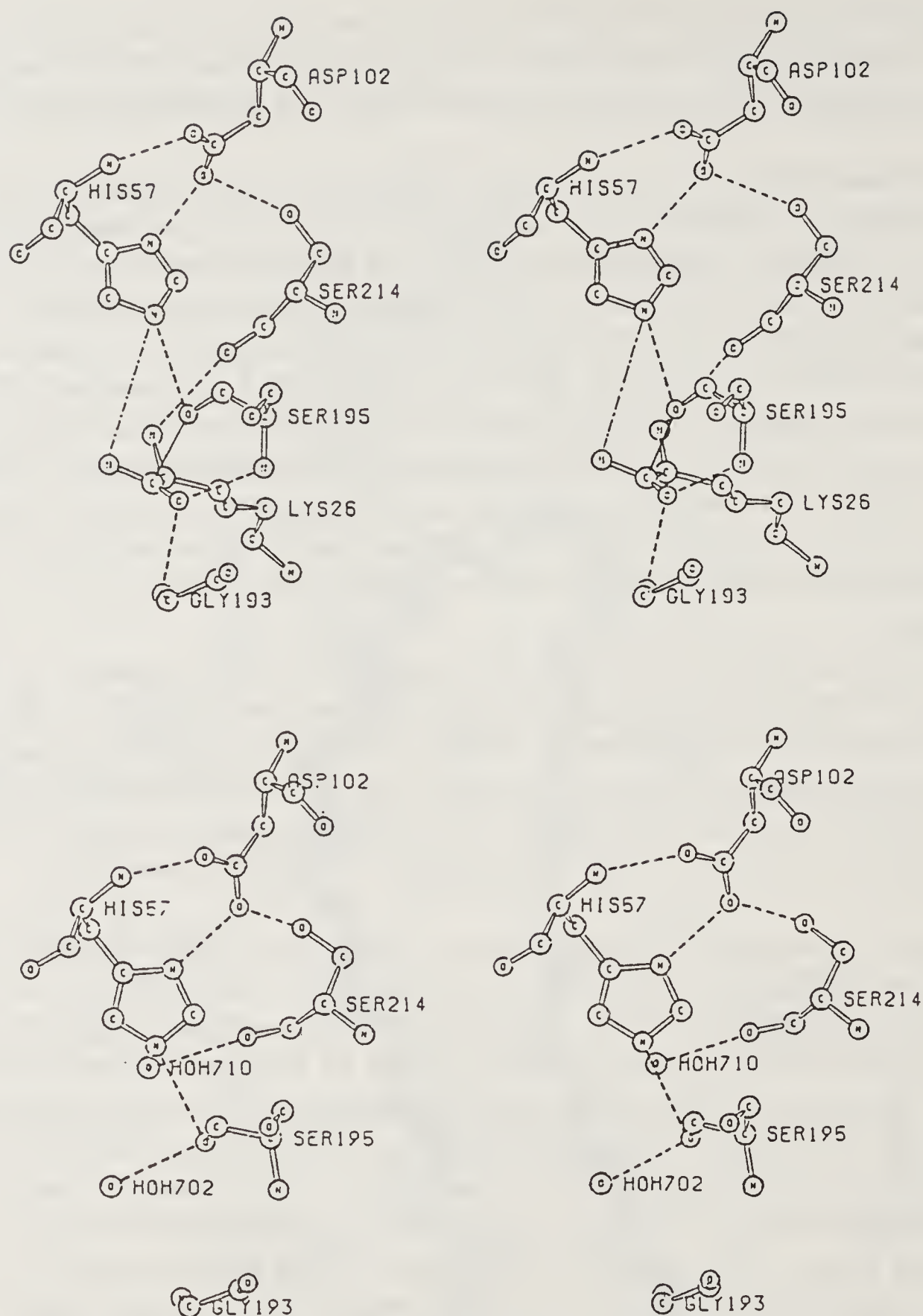


Figure 10. Stereo drawings of the active center of trypsin with (upper) and without (lower) reactive Lys 26 of AB-I.

Thr 25, N^α-Lys 31, O', and also has extensive van der Waals contacts within the ring. The rigid hydrophobic Pro-Pro sequence and the disulfide linkage also serve to stabilize the structure of the ring. This shows that the structure of the peptide chain Cys 24-Met 28 including the scissible bond, which is a "primary contact region" (or "front side") in direct contact with the enzyme, is stabilized and kept in the native form by the "secondary contact region" (or "back side") Pro 29-Cys 32 (7, 20), thus suppressing the enzymatic reaction.

Such a situation seems to be very similar in the other complexes between serine proteases and their inhibitors such as,

trypsin·BPTI (bovine pancreatic trypsin inhibitor) (21),
trypsin·STI (soybean trypsin inhibitor) (22),

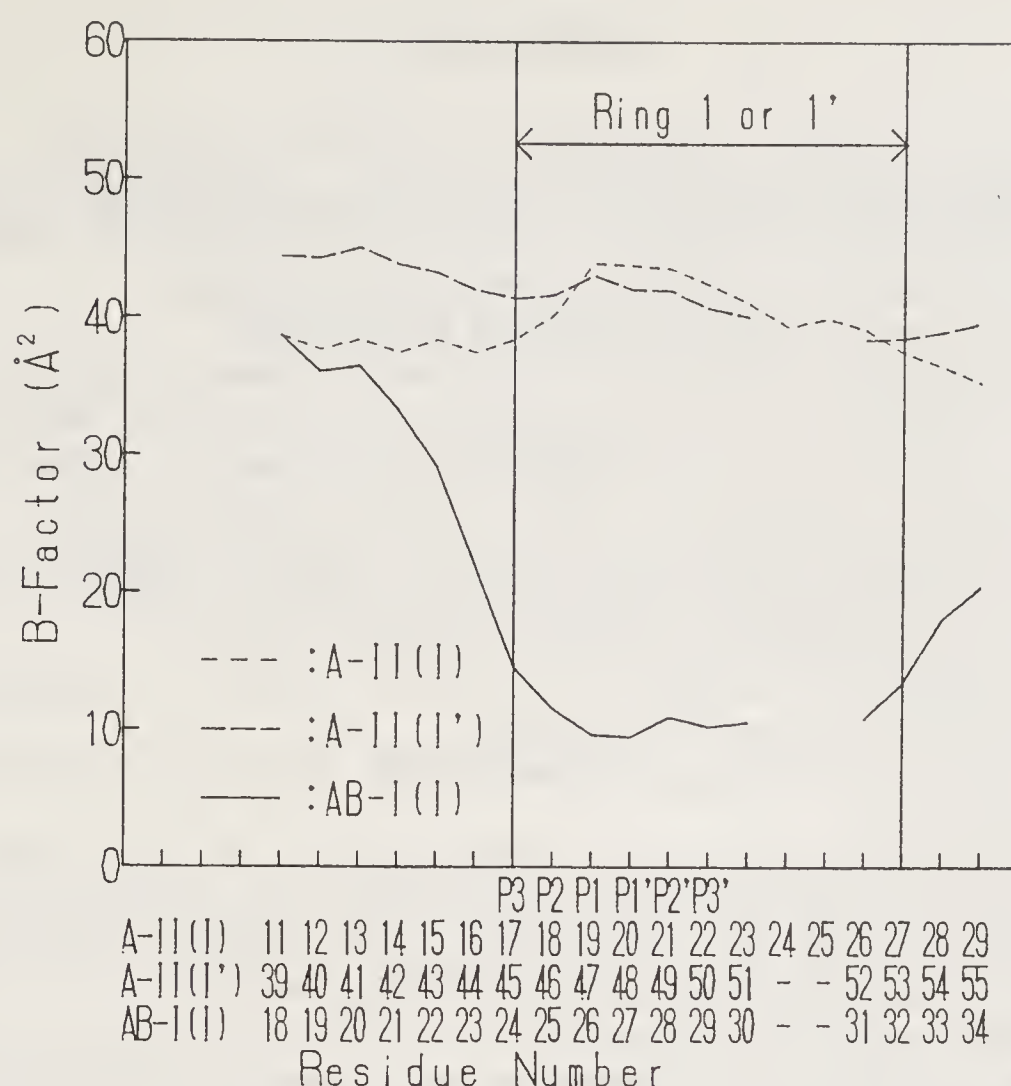


Figure 11. A comparison of the mean temperature factors of the residues in the vicinities of the reactive sites of free A-II ring 1 and 1', and those of AB-I ring 1 in the complexes with trypsin.

subtilisin BPN'·SSI (*streptomyces* subtilisin inhibitor) (20),
 trypsinogen·PSTI (pancreatic secretory trypsin inhibitor) (23),
streptomyces griseus protease B·OMTKY3 (turkey ovomucoid third domain) (24),
 chymotrypsin·OMTKY3.

The conformational angles of the reactive sites of some inhibitors in their complexes with proteases are listed in table 1. The similarity among these structures is remarkable.

Since the reactive site of the inhibitor should make an extensive close contact with the enzymatic active center, its structure should precisely be designed, thus different inhibitors may often have very similar structures of the reactive sites. The inhibitor must have a secondary contact region (or back side), however, restrictions on the structure may not be severe, since their role is only to stabilize the primary contact regions with relatively weak interactions of hydrogen bonds and van der Waals contacts except for disulfide linkages. Thus various structures may be possible for the secondary contact region. The inhibitory mechanism described in this report seems to be generally applicable in many cases.

Acknowledgements

We are grateful to Professors Tokuji Ikenaka, Hara Saburo and Kenji Takahashi and Drs. Isao Tanaka, Kazue Watanabe, Chikako Ishikawa, Nakamura Shin and Norioka Shigemi for their kind help and encouragement during this study.

References

1. Laskowski, M., Jr. & Kato, I. (1980) *Annu. Rev. Biochem.* **49**, 593–626.
2. Abe, J., Nakamura, H., Shibata, H., Norioka, S., Hara, S. & Ikenaka, T. (1985) *Jpn. Agric. Res. Quant.* **18**, 229–232.
3. Wilson, K. A. & Chen, J. C. (1983) *Plant Physiol.* **71**, 341–349.
4. Norioka, S. & Ikenaka, T. (1983) *J. Biochem.* **94**, 589–599.
5. Tanaka, Y., Tsunogae, Y., Hayakawa, Y., Tanaka, I., Yamane, T., Ashida, T., Ishikawa, C., Watanabe, K., Nakamura, S. & Takahashi, K. (1983) *J. Biochem.* **94**, 611–613.
6. Tsunogae, Y., Suzuki, A., Sone, T., Takahashi, K., Tanaka, I., Yamane, T., Ashida, T., Norioka, S., Hara, S. & Ikenaka, T. (1986) *J. Biochem.* **100**, 243–246.
7. Tsunogae, Y., Tanaka, I., Yamane, T., Kikkawa, J., Ashida, T., Ishikawa, C., Watanabe, K., Nakamura, S. & Takahashi, K. (1986) *J. Biochem.* **100**, 1637–1646.
8. Tsunogae, Y. (1987) PhD Thesis, Nagoya University.
9. Suzuki, A., Tsunogae, Y., Tanaka, I., Yamane, T., Ashida, T., Norioka, S., Hara, S. & Ikenaka, T. (1987) *J. Biochem.* **101**, 267–274.
10. Ishikawa, C., Watanabe, K., Sakata, N., Nakagaki, C., Nakamura, S. & Takahashi, K. (1985) *J. Biochem.* **97**, 55–70.
11. Norioka, S., Omichi, K. & Ikenaka, T. (1982) *J. Biochem.* **91**, 1427–1434.
12. Sakabe, N. (1983) *J. Appl. Crystallogr.* **16**, 542–547.
13. Wang, B. C. (1985) in *Methods in Enzymology*, eds. Wyckoff, H. W., Hirs, C. H. W. & Timasheff, S. N. (Academic, New York), Vol. 115, pp. 90–112.
14. Tanaka, I. (1986) unpublished program.
15. Hendrickson, W. A. (1985) in *Methods in Enzymology*, eds. Wyckoff, H. W., Hirs, C. H. W. & Timasheff, S. N. (Academic, New York), Vol. 115, pp. 252–270.
16. Suzuki, A., Yamane, T., Ashida, T., Norioka, S., Hara, S. & Ikenaka, T. (1990) manuscript in preparation.
17. Norioka, S. & Ikenaka, T. (1983) *J. Biochem.* **93**, 479–485.
18. Mitsui, Y., Satow, Y., Watanabe, Y. & Iitaka, Y. (1979) *J. Mol. Biol.* **131**, 697–724.
19. Bode, W. & Schwager, P. (1975) *J. Mol. Biol.* **98**, 693–717.
20. Mitsui, Y., Satow, Y., Watanabe, Y., Hirono, S. & Iitaka, Y. (1979) *Nature (London)* **277**, 447–452.
21. Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. & Steigemann, W. (1974) *J. Mol. Biol.* **89**, 73–101.
22. Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. (1974) *Biochemistry* **13**, 4212–4228.
23. Read, R. J., Fujinaga, M., Sielecki, A. R. & James, M. N. G. (1983) *Biochemistry*, **22**, 4420–4433.
24. Fujinaga, M., Sielecki, A. R., Read, R. J., Ardelt, W., Laskowski, M. Jr., & James, M. N. G. (1987) *J. Mol. Biol.* **195**, 397–418.

Complex Flavoenzymes – Models for Electron Transfer Complexes

F. Scott Mathews

Department of Cell Biology and Physiology, Washington University School of Medicine, St. Louis, MO 63110, USA

Introduction

Electron transport plays a key role in many metabolic processes. These processes range from the catabolic oxidation of substrates coupled to oxidative phosphorylation in mitochondria to photosynthesis in plants and bacteria. These chemical reactions are catalyzed by a variety of enzymes and electron transfer proteins which transmit electrons along a series of reaction steps, separated in free energy by relatively small differences in potential, to some terminal reaction such as the reduction of molecular oxygen to water.

In our laboratory we are studying several enzymes which catalyze the oxidation of a substrate in a 2-step process. In the first step, substrate oxidation occurs with transfer of 2 electrons to an enzyme-bound cofactor, yielding oxidation product and a 2-electron reduced cofactor. In the second step, electrons are transferred, one at a time, to a second cofactor capable of taking up only a single electron. The latter cofactor is either a heme or an iron–sulfur cluster and is located either in the same subunit as or a different subunit from the flavin. In each case electrons are transferred in a subsequent step to another electron transfer protein.

Three enzymes will be discussed in this article, trimethylamine dehydrogenase (TMADH), flavocytochrome b_2 and *p*-cresol methylhydroxylase (PCMH). The first 2 contain FMN and the third FAD as the primary electron acceptor. In TMADH, the second cofactor is a 4-iron, 4-sulfur cluster bound to the same protomer as the FMN. In flavocytochrome b_2 and PCMH, the second cofactor is a heme group. In the case of flavocytochrome b_2 the heme is part of a cytochrome domain attached to the FMN-binding protomer. In PCMH, the heme is in a separate cytochrome subunit.

The discussions of these 3 enzymes will focus on both aspects of the oxidation reduction reaction. The first aspect is the primary one of substrate oxidation and flavin reduction. The active site geometry, mode of flavin binding and many other structural features differ in the three enzymes. Comparison of these structures provides a wealth of information on the mechanistic details of the primary step.

The second aspect of importance in these studies is the 1-electron transfer step. This step is representative of a general class of reactions and occurs between a wide variety of electron transfer proteins. Fundamental to this step is formation of an electron transfer complex. Although the structures of many electron transfer proteins

are known, very few structures of complexes between such proteins have been determined. Instead, a number of theoretical models for these complexes have been proposed. Examples of such models are flavodoxin–cytochrome *c* (1), cytochrome *b*₅–cytochrome *c* (2) and cytochrome *b*₅–hemoglobin (3). The major common features in these models are that the surface topology and charge distributions of the interacting proteins are complementary and that the prosthetic groups are approximately coplanar. On the other hand, there is some variability in the distance separating prosthetic groups. For example, the minimum distance between porphyrin ring atoms in the models for the cytochrome *b*₅ complexes is about 8–10 Å whereas the heme and FMN are in van der Waals contact in the model for the flavodoxin–cytochrome *c* complex.

The present studies provide 3 specific examples of naturally occurring electron transfer complexes between a flavin group and a second redox cofactor, either a heme or iron–sulfur cluster. These studies can be used to examine the validity of the theoretical models proposed previously and to provide data for other possible complexes.

Trimethylamine Dehydrogenase

Trimethylamine dehydrogenase (TMADH) is an iron–sulfur flavoprotein which catalyzes the oxidative N-demethylation of trimethylamine to form dimethylamine and formaldehyde (4). It is found in several methylotrophic bacteria. The protein is a symmetrical dimer of *M*_r 166,000 (5, 6). Each subunit contains one [4Fe–4S] center and one FMN cofactor. The latter is bound covalently through the 6 position of the flavin ring to a cysteine side chain (7). The natural electron acceptor is an electron transferring flavoprotein of *M*_r 57,000 (8).

The mechanism of action of TMADH has been studied by absorption spectroscopy, stopped flow and freeze quenched kinetic methods (9) and by EPR spectroscopy (10). When the enzyme is reduced by excess substrate (1 to 2 mM) under anaerobic conditions, substrate binding and full reduction of the flavin to the hydroquinone state occurs within a few milliseconds. After the flavin is reduced, a triplet state is formed upon transfer of one electron from the reduced flavin to the [4Fe–4S]⁺ center. This triplet state has a very intense EPR signal which arises from the coupling of two *S* = 1/2 spin systems on the flavin semiquinone and the reduced [4Fe–4S]⁺ group to form an *S* = 1 spin system (11). Formation of the triplet state occurs much more slowly, in 300–400 ms, and may be the rate limiting step during catalysis.

When TMADH is treated with excess dithionite under anaerobic conditions, it takes up a total of three electrons per subunit to form the fully reduced flavin and reduced [4Fe–4S]⁺ cluster. The EPR signal is typical of an isolated [4Fe–4S]⁺ cluster. However, in the presence of tetramethylammonium chloride, a competitive inhibitor of TMADH, dithionite is only able to force the enzyme to take up two electrons per subunit, allowing full development of the triplet state.

Structure Analysis

TMADH from the bacterium W3A1 crystallizes (5) in space group P2₁ with cell parameters *a* = 147.6 Å, *b* = 72.0 Å, *c* = 83.7 Å and β = 97.6°. There are 2 subunits per asymmetric unit related by a non-crystallographic 2-fold axis. At the time the

structure was solved, the amino acid sequence was unknown except for that of a 12 residue peptide which was covalently attached to the flavin (7). The structure of TMADH was determined at 2.4 Å resolution (12) by the MIRAS method. Data were recorded from one crystal each of the native protein and of 2 heavy atom derivatives on a Hamlin detector system (13) at the UCSD area detector facility. Anomalous scattering data were recorded and scaled separately for all 3 crystals. Agreement *R* factors of 3% to 4% were obtained for the approximately 130,000 Friedel-independent reflections. There was on average approximately 3-fold redundancy for each observed reflection. The electron density function was averaged (14) about the molecular 2-fold axis and an alpha-carbon chain traced, first on a minimap and then on a molecular graphics system. Positions for 733 alpha-carbon atoms were obtained in this way. In addition, the positions of the flavin group and [4Fe-4S] cluster were clearly visible.

We have improved the quality of the electron density map by solvent levelling (15). Although the crystals contain 54% solvent, the solvent content parameter was set to 51% as a precaution against artificially truncating external side chains. Using this procedure the average figure of merit increased from 0.56 to 0.81 with an accumulated phase shift of 29° after 3 cycles of solvent leveling. This resulted in considerable improvement in the map. The solvent-leveled map was then averaged about the molecular 2-fold axis prior to interpretation.

An hypothetical "X-ray" sequence (16) was fitted to the solvent-leveled and averaged 2.4 Å resolution electron density map. Individual amino acids were added one at a time to the growing polypeptide chain and adjusted to fit the side chain density. Often several different amino acids were tested to achieve the closest match to the density. A total of 729 amino acids were identified and fitted to the density, compared to 733 residues previously estimated from the α -carbon fitting. In well defined regions of the map almost all of the amino acids could be easily identified by shape. In other regions, especially on surface loops, only the larger aromatic amino acids were unambiguous. In the X-ray sequence, a total of 8 cysteine and 16 methionine were identified.

As an initial check on the X-ray sequence, an anomalous difference Fourier map was computed from the native X-ray data. This map showed prominent peaks at the known iron, phosphate and cysteine positions and verified most of the methionine residues chosen. All of the eight cysteine residues occur at chemically identified locations, i.e., the five covalent linkages to the FMN and [4Fe-4S]²⁺ cofactors and the three mercury binding sites utilized for phase determination.

The amino acid sequence of about 65% of the residues of TMADH have been determined using gas phase sequencing methods (M. Barber, personal communication). The X-ray sequence was used to align peptide fragments. Approximately 75% of the "X-ray" sequence was found to be correctly identified.

We have carried out preliminary least-squares refinement of TMADH at 2.4 Å resolution using the previously-determined X-ray sequence. Ten cycles of Hendrickson-Konnert (17) refinement were carried out using a single overall temperature factor. No solvent has been added. For the final cycle, using data between 5.0 and 2.4 Å resolution, the *R* factor was 0.226 with an r.m.s. deviation from ideal bond lengths of 0.023 Å.

Structure of the Molecule

The TMADH dimer is an ellipsoidal disk approximately 100 Å in diameter and 65 Å thick. The dimeric molecule is shown in figure 1. The folding topology of the

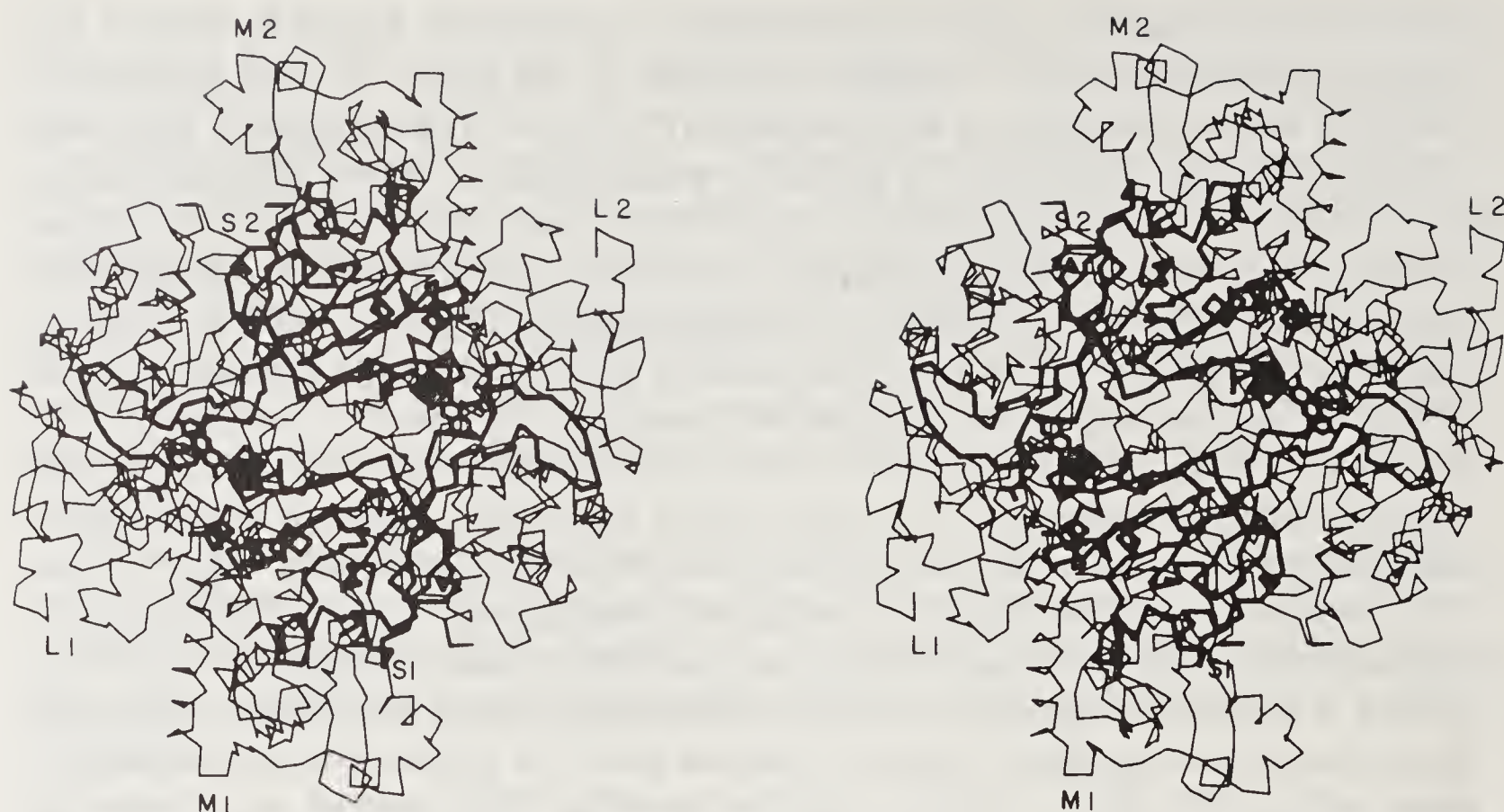


Figure 1. Dimer of TMADH. The molecule is viewed along the molecular 2-fold axis. The large, medium and small domains of each subunit are labeled L1, M1, S1, L2, M2 and S2, respectively. The small domains and the [4Fe-4S] and FMN prosthetic groups are shown in heavy lines.

TMADH subunit is shown in figure 2. Each subunit of TMADH consists of 3 domains. The largest domain is made up of approximately the first 380 residues starting from the N-terminus. It is folded into a parallel $\beta_8\alpha_8$ barrel and contains both the FMN and the $[4\text{Fe}-4\text{S}]^{2+}$ cofactors (figure 3). The FMN is located at the C-terminal end of the β barrel, with atom C-6 of the flavin ring bound to a cysteine sulfur atom at position 30, near the end of the first β strand. The O-2 atom of the flavin ring is nearly coincident with the barrel axis.

The phosphate group of FMN is located between the ends of strands 7 and 8 close to the N-terminus of helix D (figure 2) just before helix 8 of the barrel. In this position the negative charge of the phosphate group can be partially compensated by the positive helix dipole (18). The flavin ring is completely buried, but located at the bottom of a channel about 14 Å deep with the *si* face of the isoalloxazine ring open to the channel. This channel opens into the center of the dimer and thence to another channel defined by the molecular 2-fold axis separating the two small domains (figure 1).

The $[4\text{Fe}-4\text{S}]^{2+}$ is bound to an α/β hairpin connected to the C-terminus of the eighth helix of the $\beta_8\alpha_8$ barrel. This hairpin lies on the external surface of the $\beta_8\alpha_8$ barrel and partially covers the benzenoid ring of the flavin. However, the iron-sulfur cluster is deeply buried in the protein, surrounded by the second and third domains. The relative positions of the FMN and $[4\text{Fe}-4\text{S}]$ cluster are shown in figure 4. The centers of the two prosthetic groups are about 12 Å apart. However, the 8 α -methyl of FMN is about 6 Å from the closest iron atom and only about 4 Å from the cysteinyl sulfur atom to which this iron is bonded. Thus the two prosthetic groups appear to be in van der Waals contact. This 4 Å distance is consistent with the distance deduced from the e.p.r. and kinetic studies in solution (10) and accounts for the strong spin coupling observed between these groups (11).

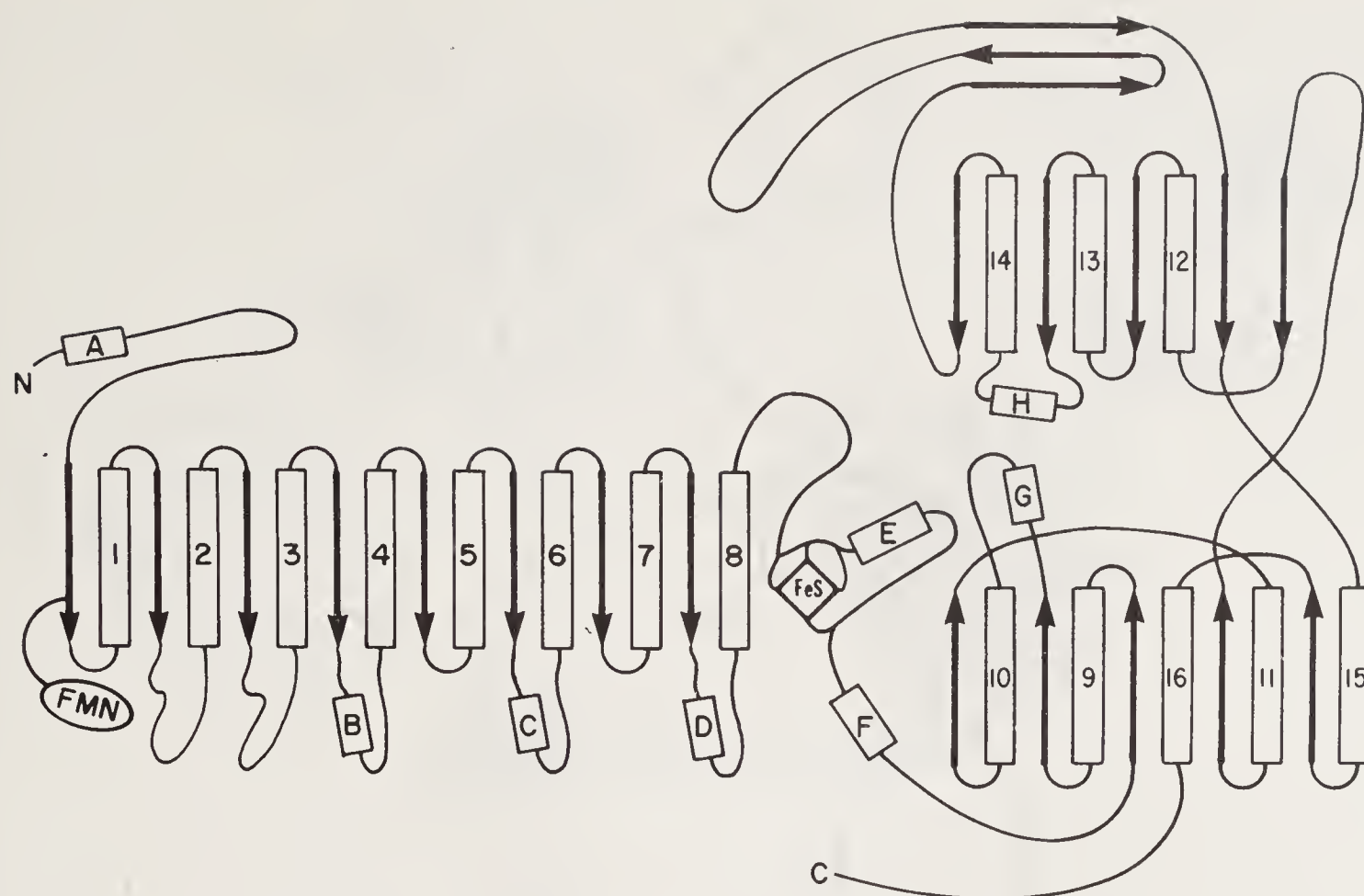


Figure 2. Schematic diagram showing the topology and secondary structure elements of TMADH. Solid arrows represent beta strands and open rectangles represent helices. Helices contained within elements of α/β supersecondary structure are numbered consecutively while helices external to these elements are designated by letters. The approximate points of attachment of the FMN and the Fe_4S_4 cluster are indicated.

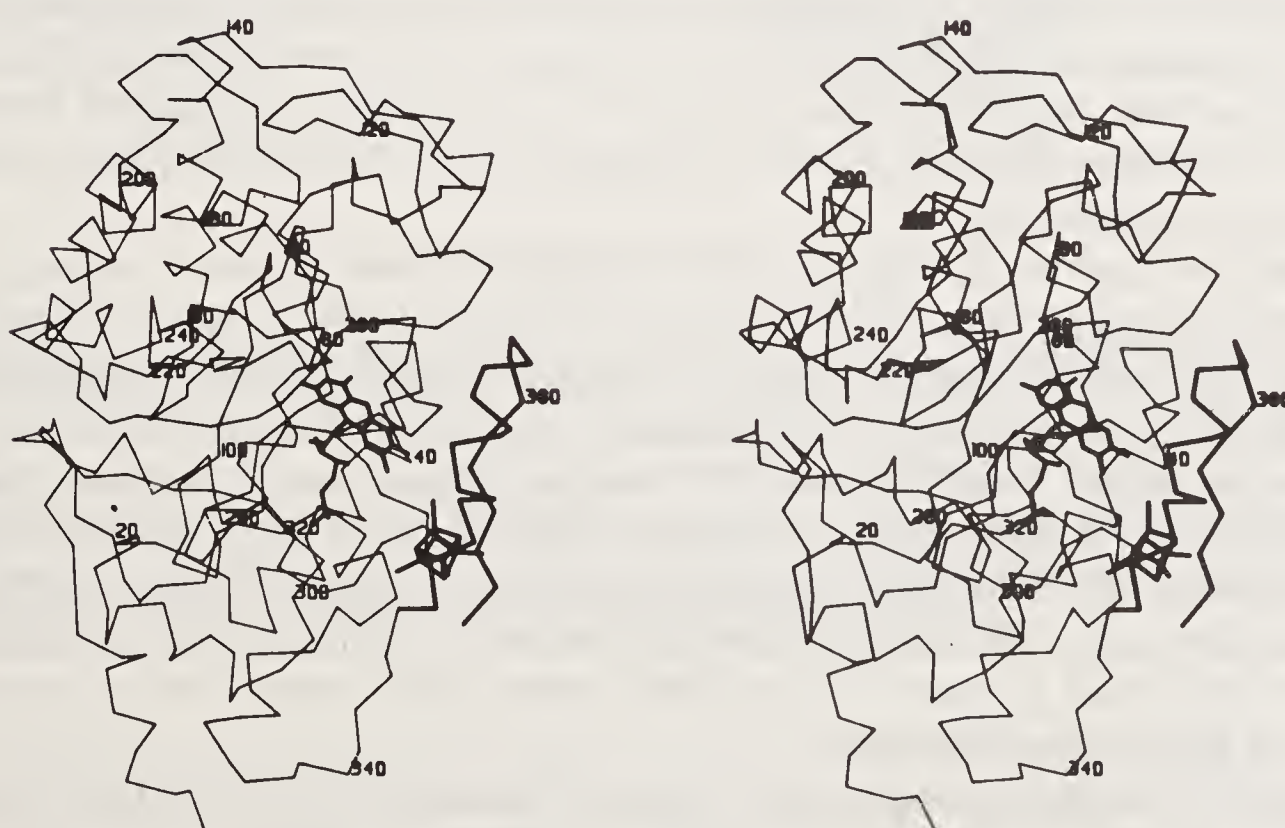


Figure 3. Stereo view of the α -carbon backbone of the N-terminal domain of TMADH. Residues 1–366 are included with every 20th residue labeled. The FMN and $[\text{4Fe-4S}]$ cofactors are shown in heavy lines. The FMN is bound at the C-terminal end of a parallel β_8 structure. The $[\text{4Fe-4S}]$ cluster is bound to an α/β loop, also shown in heavy lines, located after the eighth α -helix of the $\beta_8\alpha_8$ barrel.

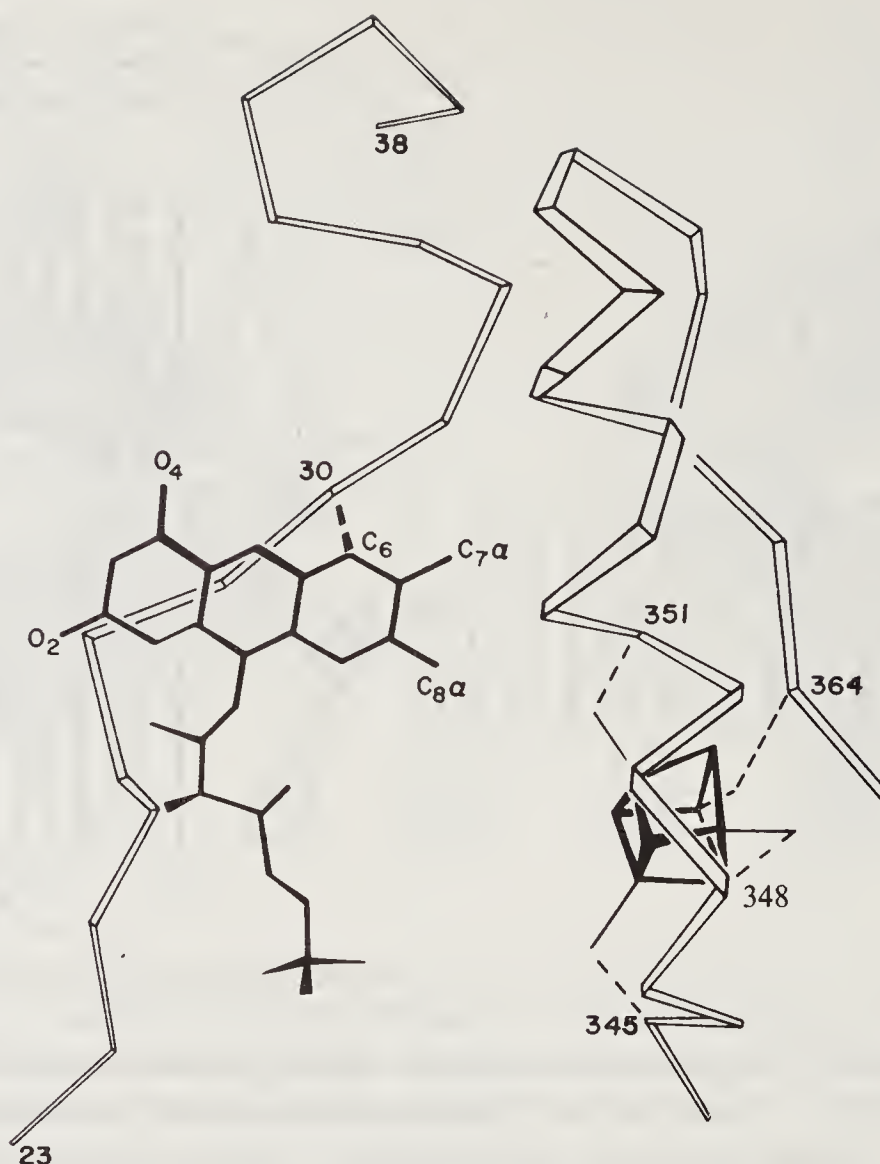
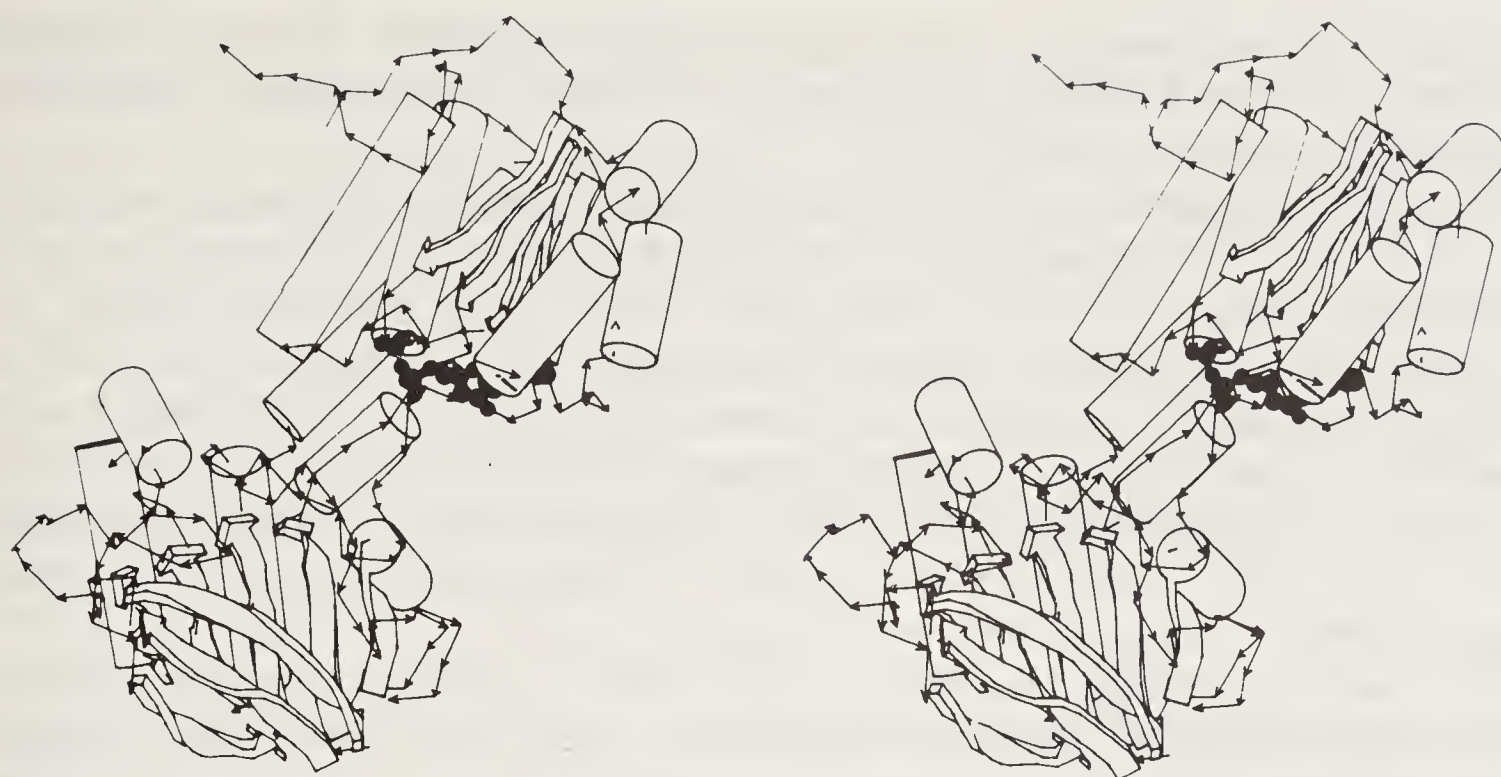


Figure 4. Diagram showing the relative orientations of the FMN and [4Fe-4S] cluster and portions of the polypeptide chains to which they are attached.

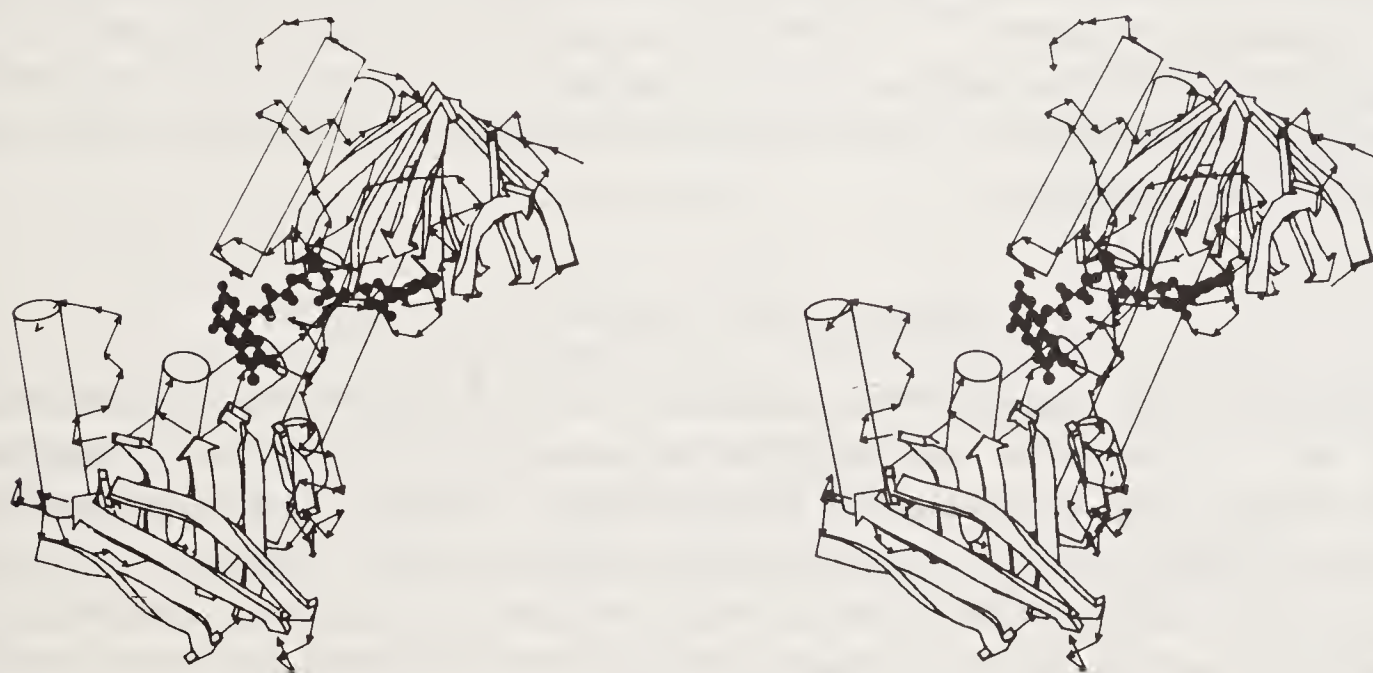
The second (medium) and third (small) domains each contain 5-stranded parallel beta sheets (figure 2). In the medium domain the central β -sheet is flanked by two or three α -helices on each side. This domain resembles the nucleotide-binding domain of lactate dehydrogenase (19) in its folding topology. In the small domain, the β -sheet is flanked by three helices on one side and by a 3-stranded antiparallel β -sheet on the other. Its folding topology closely resembles that of the FAD-binding domain of glutathione reductase (20).

The medium and small domains are connected at two places. The polypeptide chain, after leaving the large domain, forms about 3/4 of the medium domain, goes on to form the small domain and finally re-enters and completes the medium domain. The topological connectivity of the medium and small domains *together* closely resembles that of the NADPH and FAD binding domains of glutathione reductase (20) (figure 5). When the α -carbon positions of both pairs of domains are compared, a large fraction of the secondary structural elements are found to nearly superimpose. The central β -strands, the strands connecting the pairs of domains and the antiparallel β -sheet of the small domain of TMADH match quite closely the corresponding elements in glutathione reductase.

In TMADH, the flavin ring is clearly bent. The electron density of the flavin ring can be most easily matched by introducing a 20° "butterfly" bend about an axis defined by the N-5 to N-10 bond. Refinement of the protein coordinates based on the "X-ray" sequence has confirmed this bent conformation. This bend angle is considerably larger than observed in other flavoproteins and approaches that observed in crystal structures of some model flavin compounds (21).



TMADH Medium and Small Domains



Glutathione Reductase Res 18-360

Figure 5. (a) Stereoscopic diagram of the two carboxy-terminal domains of trimethylamine dehydrogenase. Ribbons represent beta strands and cylinders represent helices. The ADP cofactor, solid circles, is attached to the upper domain with the pyrophosphate moiety at the left. (b) Stereoscopic diagram of the FAD-binding (upper) and NADPH-binding (lower) domains of glutathione reductase. The FAD cofactor is also shown, as solid circles. Pictures produced by a computer program written by A. M. Lesk and K. D. Hardman.

During the fitting of the hypothetical “X-ray” sequence to the electron density, an unconnected piece of density was discovered which previously had been interpreted as part of the polypeptide chain. This density corresponded in position to the ADP moiety of FAD in the analogous position in glutathione reductase. The location of ADP in the medium and small domains is also included in figure 5 and can be compared with the position of FAD in glutathione reductase. The anomalous difference Fourier also contained peaks of high density at the diphosphate position

of magnitude comparable to the single phosphate of FMN. Thus, the structural similarity between TMADH and glutathione reductase now includes the presence of the ADP fragment of FAD.

The functional significance of ADP in TMADH is unknown. There is no electron density beyond the diphosphate portion of the ADP which would correspond to the riboflavin portion of FAD. In fact, that region of the map corresponds to a solvent-filled cavity near the surface of the protein. One possibility is that ADP is a vestigial fragment from a glutathione reductase-like precursor of TMADH. Another possibility is that FAD was originally present in TMADH but was cleaved during isolation. Modeling studies suggest that if FAD were present, it could lie quite close to the $[4\text{Fe-4S}]^{2+}$ and serve as a conduit for electrons to the electron transfer flavoprotein.

The mode of binding of FMN in TMADH is remarkably similar to that in flavocytochrome b_2 (see ref. 30) and glycolate oxidase (see ref. 41). The flavin-binding domain in each of these enzymes is made up of a parallel $\beta_8\alpha_8$ barrel structure. In all three structures, the FMN cofactors nearly coincide when the β barrels are optimally aligned. Thus, the covalent nature of the binding of FMN to TMADH does not result in a mode of binding significantly different from that in other, non-covalently bound FMN containing enzymes. The identification of more subtle differences in FMN binding between the covalently and non-covalently bound cases, such as alteration in the degree of bending or in the electronic structures of the flavin ring requires continued investigation.

Ligand Binding and Oxidation State Changes

Difference Fourier studies were carried out at 6.0 Å resolution (22) from crystals soaked with the inhibitor tetramethylammonium chloride (TMAC), the substrate trimethylamine (TMA) and with the reducing agent dithionite both alone and in the presence of TMAC. These conditions allowed the crystalline enzyme to be studied in the oxidized, fully reduced and triplet states, and the binding of substrate and inhibitor to be compared. In all soaks, the significant difference peaks were confined for the most part to within 12 Å of the flavin ring, either within the active site pocket or close to nearby side chains.

TMAC bound to the oxidized enzyme produced the fewest structural changes. The largest peak occurred above the flavin ring opposite N5 of the flavin ring and probably represents the inhibitor binding site. Both TMA, and TMAC with dithionite, which cause formation of the triplet state of the enzyme, gave rise to greater apparent changes in structure. The largest peak still occurred opposite the flavin N5 position and probably represents bound substrate or product. Other difference density occurred in the active site pocket, near the helix linked to the iron-sulfur cluster or close to the benzenoid portion of the flavin. In neither case is there any apparent movement of the flavin ring or iron-sulfur cluster as a whole, nor any other large conformational changes. Dithionite alone, which produces the fully reduced singlet state, showed the greatest changes in the difference map. The largest peak again occurred above the flavin ring, but was slightly displaced. There were also larger positive and negative peaks near the benzenoid portion of the flavin ring.

The studies of TMAC and of dithionite binding alone have been extended to 2.8 Å resolution and 2.4 Å resolution, respectively. They both provide considerably greater

detail of the binding of inhibitor and of reduction. In each case data were recorded on a Hamlin detector system from crystals soaked in 25 mM TMAC or dithionite for about 15 minutes prior to mounting in a capillary filled with the appropriate artificial mother liquor. The data were recorded with about 3-fold redundancy and gave agreement *R* factors of about 8% and 5% respectively.

In the TMAC difference map at 2.8 Å resolution, there is a major peak at the active site above the N5 position of the flavin ring, in a similar position as found in the low resolution study. There is also a positive and negative pair of peaks near Trp 355 indicating side chain movement in the plane of the ring away from the inhibitor binding site. Trp 264 also moves slightly in a direction perpendicular to the ring plane, away from the ligand binding site.

Dithionite alone shows no ligand binding in the 2.4 Å resolution difference map. However, Trp 355 shows the same movement away from the ligand binding site. Trp 264, on the other hand, shows a very large movement, this time in a direction perpendicular to the ring plane in the opposite direction to that found with TMAC. It is significant that in both the TMAC derivative and the dithionite reduced crystals, there is no difference electron density near the flavin ring. There are a number of other positive and negative peaks at the active site which cannot be interpreted yet since the structure is not fully refined.

The difference Fourier studies described above are interpreted on the assumption that the enzyme in the crystalline state undergoes the same changes in oxidation state in the presence of substrate or reducing agent as the enzyme in solution. To verify these assumptions, optical spectroscopic and ESR studies of TMADH in the crystalline state were carried out under conditions similar to those of the difference Fourier studies. The optical spectra were measured in collaboration with Dr. Marvin Mäkinen of the University of Chicago and the ESR studies were done in collaboration with Dr. Richard Dunham of the University of Michigan.

The ESR and spectroscopic studies of TMADH crystals show that under the conditions of the crystalline soaking experiments, the redox states of the enzyme are the same as would occur in solution. Thus crystal packing forces do not inhibit formation of the reduced or triplet states of the enzyme. Any structural changes which occur in the crystals under these conditions probably occur in the enzyme in solution.

Flavocytochrome b_2

Flavocytochrome b_2 (L-lactate: ferricytochrome c oxidoreductase) is a tetrameric enzyme of M_r 230,000 located in the intermembrane space of yeast mitochondria. It contains two cofactors, FMN and heme, both bound noncovalently (23). The enzyme catalyzes the oxidation of lactate to pyruvate with subsequent transfer of electrons to cytochrome c. Substrate oxidation occurs *via* transfer of two electrons from lactate to FMN to form the flavin hydroquinone. The flavin is then reoxidized in two 1-electron transfer steps to the endogenous heme group with formation of a semiquinone intermediate. *In vivo*, flavin reoxidation is coupled to electron transfer from the heme group to cytochrome c (24).

The amino acid sequence of the enzyme is known (25). Residues 8–103 can be released from the enzyme as a stable entity by trypsin digestion. This fragment, called cytochrome b_2 core, is homologous to the soluble fragment of microsomal cytochrome

b₅ (26). The remaining 408 residues contain the flavin-binding domain. Included in this piece is a short region, near residue 310, which is susceptible to digestion by yeast or other proteases in the absence of protease inhibitors such as phenylmethylsulfonyl fluoride (27). Cleavage does not result in disruption of the quaternary structure of the enzyme but does lead to an alteration of its kinetic properties (28).

Structure Analysis

The structure of flavocytochrome b₂ was determined initially at 3.0 Å resolution with X-ray diffraction data collected from native and two heavy atom derivatives (29). Crystals of flavocytochrome b₂ contain two subunits per asymmetric unit and have about 68% solvent content. They diffract to a maximum resolution of 2.4 Å. The data were collected at the area detector facility at the University of California at San Diego (UCSD) (13). The agreement *R* factor ranged from about 8% for the native data to about 12% for the derivative data and were recorded with approximately 10-fold redundancy. Heavy atom positions were obtained by difference Patterson methods and phases were calculated by the MIR method with anomalous scattering. The 3.0 Å map was improved by solvent leveling procedures (15) and by averaging the two images of the molecule contained in the asymmetric unit (14). After initially tracing the polypeptide chain in a minimap the full sequence was fitted to the electron density on a molecular graphics system.

Data were also collected to 2.4 Å resolution from native crystals by oscillation photography using synchrotron radiation at the Stanford Synchrotron Radiation Laboratory (SSRL) (30). A total of 14 crystals from a single batch of enzyme were used to record 701,250 reflections on 296 film packs to yield 68,896 unique reflections. These data were combined with data recorded between 10.0 Å and 3.3 Å from a single crystal, also from the same batch, on the area detector at SSRL (31) using synchrotron radiation tuned to near the iron absorption edge.

The model of flavocytochrome b₂ was refined using the Hendrickson–Konnert restrained refinement procedure (17). A total of 81 cycles of refinement were carried out, initially at 3.0 Å resolution using the UCSD data and subsequently at 2.4 Å refinement using only the SSRL data. Isotropic temperature factors were allowed to vary during the last 35 cycles. The two subunits were refined independently. The model was rebuilt a total of 5 times using maps calculated with $(2F_o - F_c)$ coefficients. A total of 283 water molecules were included in the final model. The final *R* factor was 0.188 and the rms deviation from ideal bond lengths was 0.014 Å.

Structure of the Molecule

The tetramer of flavocytochrome b₂ is shown in figure 6. The four subunits are related to one another by a molecular four-fold axis of symmetry. This symmetry axis is coincident with a crystallographic 2-fold axis. Each subunit is composed of two domains, a cytochrome domain, residues 1–99, and a flavin-binding domain, residues 100–486. In addition, there is a C terminal tail, residues 487–511, which extends into the centre of the tetramer, wraps around the four-fold axis and makes contact with each of the other 3 subunits.

The four flavin-binding domains pack together about the 4-fold axis to form an ellipsoidal disk about 100 Å in diameter and about 60 Å thick. The four cytochrome

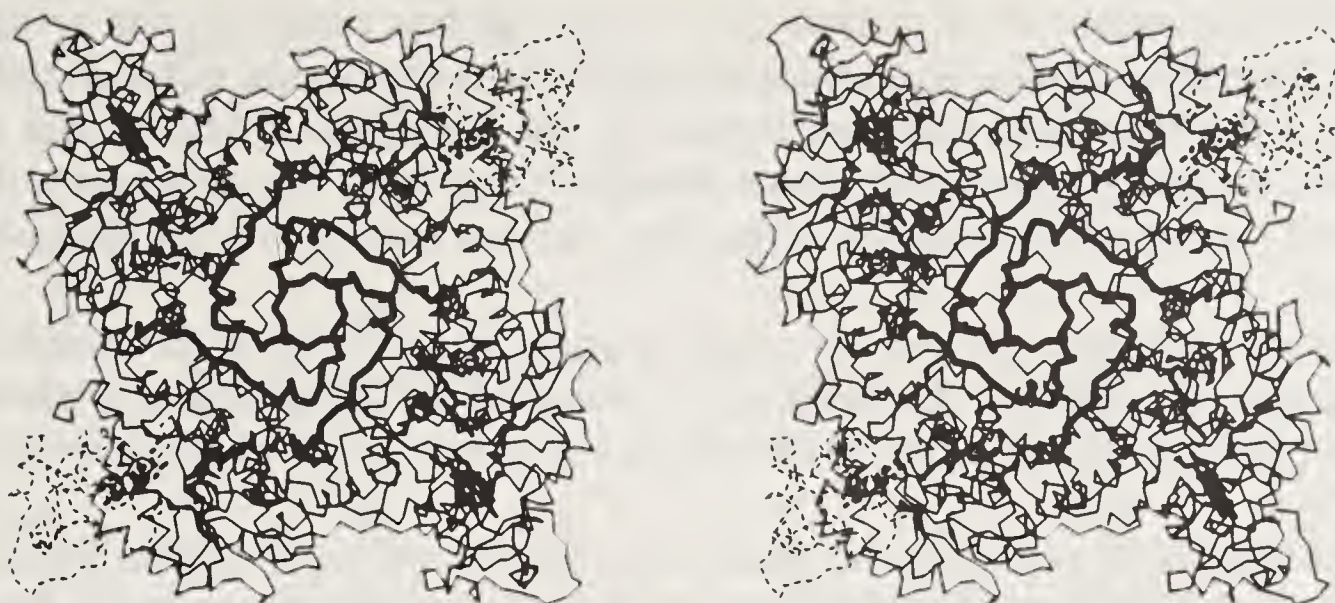


Figure 6. Stereoscopic diagram of the flavocytochrome b_2 tetramer viewed down the molecular 4-fold axis. The two disordered cytochrome domains are drawn in dashed lines while the C-terminal tail near the center of the tetramer, residues 487–511, is drawn in heavy bold lines. The heme and FMN groups are in intermediate bold lines.

domains are located on the periphery of the disk and protrude from above the midplane of the disk.

In the electron density map, only two of the four cytochrome domains are visible. The other two cytochrome domains are positionally disordered and are completely absent from the map. On the other hand, the two independent flavin-binding domains are quite similar in structure. In both subunits, residues from positions 100 to 299 and from 317 to 511 are ordered. Residues 300 to 311 are disordered in both subunits while residues 312 to 316 are disordered in subunit 1 and ordered in subunit 2. When 391 α -carbon coordinates of the two subunits are optimally aligned the r.m.s. deviation in positions is 0.36 Å.

The ordered monomer of flavocytochrome b_2 is shown schematically in figure 7. The heme group is located at the interface between the cytochrome domain and the



Figure 7. Stereoscopic diagram of the ordered subunit of flavocytochrome b_2 . The ordered cytochrome domain is below and the flavin-binding domain is above. Both the heme and flavin (solid circles) are buried. Picture produced by a computer program written by A. M. Lesk and K. D. Hardman.

flavin-binding domain. It is tucked into a hydrophobic crevice in the cytochrome domain formed by 2 pairs of antiparallel helices and a 6-stranded mixed β -sheet. The ligands of the heme iron atom are formed by the N ϵ atoms of His 43 and His 66. The heme propionate groups are both extended away from the cytochrome domain toward the centre of the molecule where they interact mainly with the flavin-binding domain.

The heme and flavin groups are nearly coplanar (figure 8). The distance from the iron atom to the center of the flavin ring is about 16 Å. The minimum separation of the two planes, from N5 of the flavin ring to the nearest atom of the pyrrole ring, is about 10 Å.

The contact surface between the cytochrome domain and the flavin-binding domain consist of both hydrophilic and hydrophobic interactions. There are 6 hydrogen bonds between the 2 domains and one direct salt bridge, from a heme propionate to Lys 296. There are also 5 water molecules bridging the 2 domains. In addition to the salt bridge, the heme propionate groups form direct hydrogen bonds to 2 tyrosine residues and bridging interactions *via* water molecules to an aspartate, an arginine and the flavin ring. The hydrophobic interactions involve 9 or 10 non-polar residues on each domain.

The flavin-binding domain is composed of a parallel $\beta_8 \alpha_8$ barrel motif. This barrel is formed by residues 189–465. The FMN is located at the C terminal end of the central β -barrel (figure 9). The isoalloxazine ring is tilted so that its plane is roughly parallel to β -strand 1, with its *si* face lying against the strand. The pyrimidine ring of the flavin lies close to the barrel axis while the benzenoid ring lies near the periphery of the β -barrel. The plane of the flavin ring is slightly bent, with a bend angle about the N5–N10 bond of about 8°. The ribityl chain is extended and lies in a groove between β -strands 7 and 8. The phosphate group is located at the N terminus of a short helix at the end of β -strand 8.

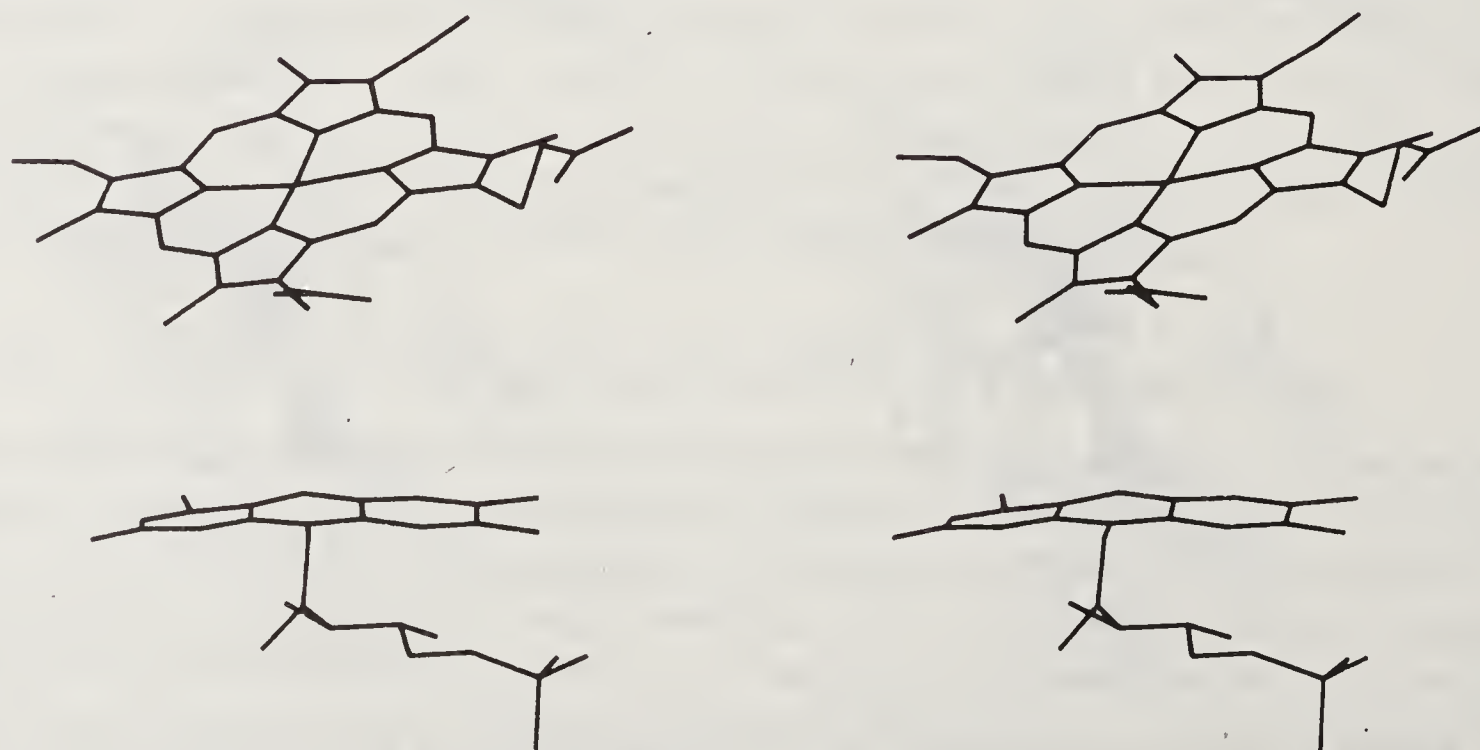


Figure 8. Stereoscopic view of the isolated heme and flavin groups of flavocytochrome b_2 .

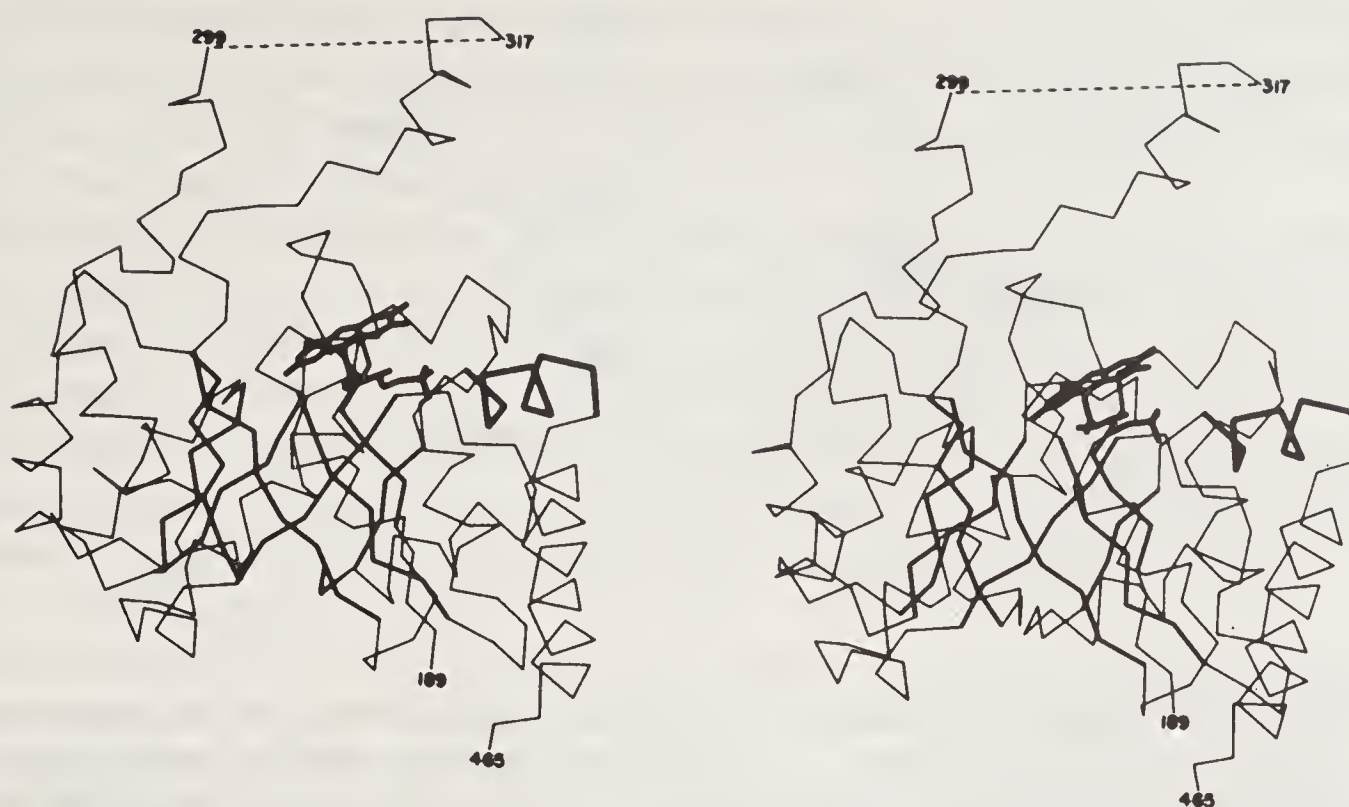


Figure 9. Stereoscopic C_α diagram of the $\beta_8\alpha_8$ barrel of the flavin binding domain of flavocytochrome b_2 . The beginning and end points of the barrel as well as the disordered loop are indicated.

Ligand Binding

In one subunit, an isolated piece of density was found close to the FMN in the unaveraged 3.0 \AA resolution electron density map. This subunit contains the disordered cytochrome domain. The density is suggestive of a bound ligand and is about the size of a molecule of lactate or pyruvate. Its center is located near the N-5 containing edge of the flavin ring about 4 \AA below the plane of the ring. The ligand is likely to be pyruvate since pyruvate has a higher affinity for the reduced enzyme than does lactate (32).

The assignment of the isolated density to pyruvate has been supported by refinement at 2.4 \AA resolution and by biophysical studies of the crystalline enzyme. For several cycles of refinement, the pyruvate molecule of subunit 2 was not restrained to be planar and the dihedral angle of the central bond, between C1 and C2, converged to a value about 20° from planarity. The group was restrained to be planar for the remainder of the refinement. Attempts to refine a molecule of pyruvate in subunit 1 were unsuccessful and the difference electron density in the corresponding location was best explained as a cluster of water molecules.

The mother liquor of the crystals grown originally in 47 mM D,L-lactate were found to contain about 2 mM pyruvate . This had probably arisen from slow turnover of the crystalline enzyme over several months with oxygen acting as the electron acceptor. In fact, pyruvate appears to stabilize the semiquinone form of the enzyme by modulation of the redox potential of the flavin group. Preliminary electron spin resonance experiments indicate that crystals of the *Saccharomyces* enzyme contain a significant signal from the flavin semiquinone radical (unpublished data). These results are consistent with pyruvate as the ligand and with half of the flavin being in the semiquinone form.

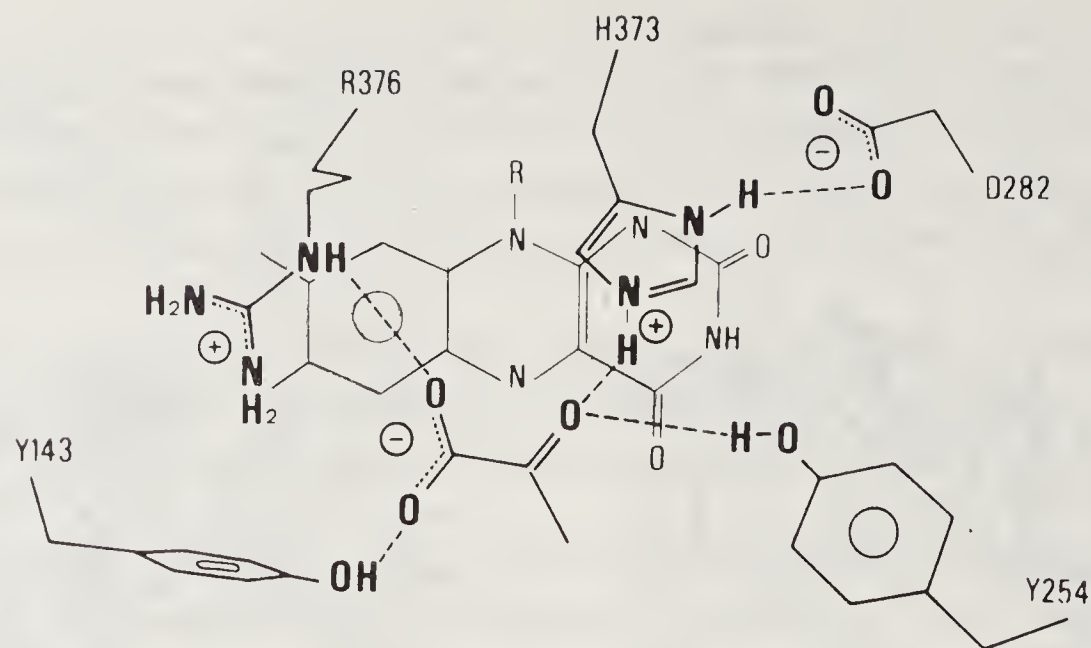


Figure 10. Schematic diagram of the chemical surroundings of the pyruvate molecule in flavocytochrome b_2 . Side chain hydrogen binding and/or ionic interactions are indicated. The flavin ring, drawn with thinner lines, lies below the plane of the pyruvate.

The chemical environment of the pyruvate is shown in figure 10. The plane of the pyruvate molecule is approximately parallel to that of the flavin ring. It is oriented so that one of the two carboxyl oxygen atoms can form a hydrogen bond with the hydroxyl of Tyr 143 while the other oxygen can form a hydrogen bond with NE of Arg 376. The buried positive charge of the latter group thus helps stabilize the negative charge of the carboxylate anion.

The keto oxygen of pyruvate interacts both with Tyr 254 OH and His 373 NE2. This histidine residue also interacts with the carboxylate of Asp 282 through the NE1 position. This latter interaction appears to help orient His 373 to bind to pyruvate.

A large body of chemical evidence favors a carbanion mechanism for flavocytochrome b_2 (33) and for several other functionally similar enzymes (34). A possible role in the catalytic mechanism of several of the amino acids which are found to interact with pyruvate, have been proposed. Lys 349 and Tyr 143 would stabilize binding of the lactate carboxylate through ionic and hydrogen bonding as in the case of pyruvate. The substrate hydroxyl could form a hydrogen bond to the hydroxyl oxygen of Tyr 254. His 373 would then be in position for its epsilon nitrogen to abstract the α -proton of the substrate (figure 10). The resulting imidazolium ion would be stabilized by interaction of the delta nitrogen with Asp 282. Since the α -carbon of lactate is tetrahedral, these last two steps would require small movements of the Tyr 254 and His 373 side chains relative to the binding of the planar pyruvate molecule.

Binding of Sulfite

The sulfite anion is a strong competitive inhibitor of flavocytochrome b_2 with an inhibition constant of $1.4 \mu\text{M}$ and a dissociation constant of $1.5 \mu\text{M}$ as measured by difference spectroscopy (35). In this respect, flavocytochrome b_2 is similar to the flavin dependent oxidases and different from most other electron transferases (36).

Differential microspectrophotometry has shown that diffusion of sulfite into preformed crystals of flavocytochrome b_2 in the oxidized state produces the same spectral perturbation as in solution (37). We have carried out a difference Fourier

study of flavocytochrome b_2 with sulfite diffused into the crystal lattice. To minimize cracking of the crystals, sulfite binding was achieved by dialysis of crystals first in lactate-free buffer and then in the presence of increasing concentration of reagent to a final concentration of $500\ \mu\text{M Na}_2\text{SO}_3$.

X-ray data were collected to $3.0\ \text{\AA}$ resolution from the sulfite-soaked crystal on a Nicolet area detector at the Argonne National Laboratory. A total of 148,300 observations of 35,800 reflections gave an R factor of 12% on intensity. The two highest peaks in the electron density difference map, about twice the highest background level, are located close to the flavin rings in the two monomers of the asymmetric unit. Each peak is located on the *si* face of the flavin and is close to the N5 atom. A sulfite anion was modeled into each of the pieces of density with the sulfur atom displaced from the N5 atom by $1.7\ \text{\AA}$ the sum of the van der Waals radii of S and N. The groups were oriented with the 3 tetrahedrally arranged oxygen atoms pointing away from the flavin ring and forming a maximum number of interactions with polar side chain atoms.

In the subunit without pyruvate (with the ordered cytochrome domain) the sulfite displaces 3 ordered water molecules from the active site. In the subunit containing pyruvate, a third peak is visible at the active site. It is a volume of negative electron density about 20% above the highest background level, having roughly the dimension of a pyruvate molecule and visible in the difference map at the pyruvate binding site (figure 11). Thus, the sulfite in the second subunit is displacing an ordered pyruvate ligand rather than water.

The simplest chemical interpretation of these observations is consistent with the mechanism of inhibition established in solution for Na_2SO_3 (38). SO_3^{2-} is the reactive species that interacts with the flavin group at the level of N5 and forms with this atom a covalent adduct. In the reaction, the pyruvate molecule, present at the active site, is displaced; in fact the affinity of flavocytochrome b_2 for sulfite determined in solution is at least 200 fold higher than that for pyruvate (35).



Figure 11. Stereoscopic view of the sulfite difference density superimposed on the active site of subunit 2 of flavocytochrome b_2 . The electron density peak on the right is positive (contour level $+4.0$ arbitrary units) and the one on the left is negative (contour level -2.5 arbitrary units). The pyruvate molecule displaced by sulfite is not shown.

The protein environment at the active site near the flavin N5 can easily explain the high affinity of the flavin for sulfite. First, the sulfite anion appears to be bound covalently to the flavin since both of the difference density peaks are located at the expected covalent distance of 1.7 \AA from N5. One of the oxygen atoms can form hydrogen bonds to Tyr 254 and His 373. A second oxygen lies close enough to Arg 376 to form a charge interaction, partially neutralizing the negative charge on the sulfite anion. The third oxygen may form a weak hydrogen bond with Tyr 143. All of these protein side chains are implicated in pyruvate binding and may be involved in the catalytic process (33). The only difference in the environment of the two sulfite ligands is the close approach of the heme propionate oxygen in subunit 1 (absent in subunit 2) and the possible interaction with a water molecule in that subunit.

Comparison with other Proteins

The cytochrome domain is very similar in conformation to microsomal cytochrome b_5 (39). The two structures are compared in figure 12. The sequence identity is very high (about 37%) for about the first 63 residues (26). Beyond this point, one deletion and two insertions are needed in the b_2 sequence to maintain maximal but reduced sequence similarity. The β -sheet and two of the four helices which form the heme crevice (in the back in figure 12) are oriented very similarly in the two structures. On the other hand, helix 4 and, to some extent helix 3 are oriented somewhat differently and reflect considerable difference in local conformation of the polypeptide chain. This results in the heme group of flavocytochrome b_2 appearing to be more exposed

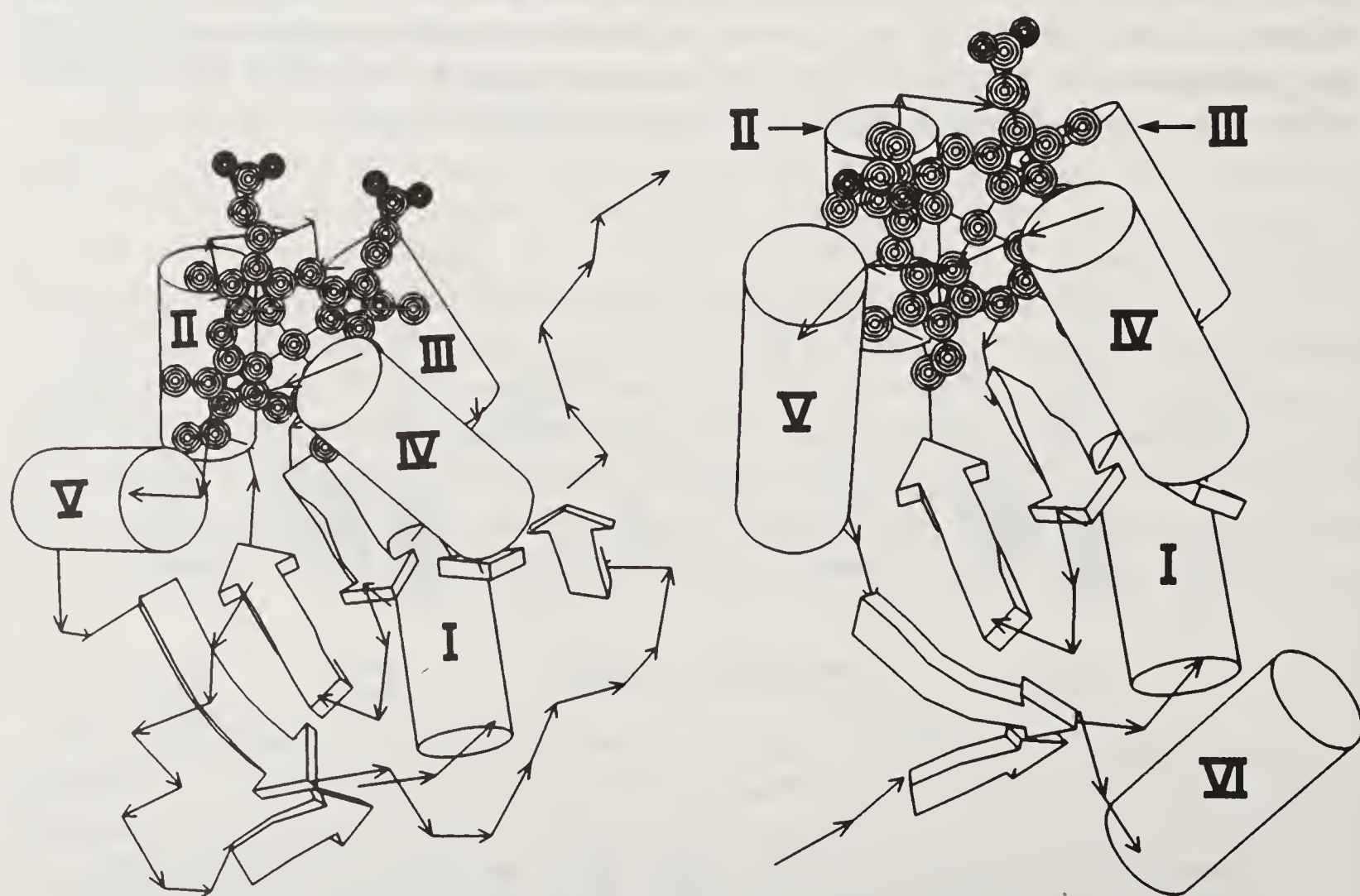


Figure 12. Schematic diagrams of (a) the flavocytochrome b_2 core domain and (b) cytochrome b_5 . Cylinders representing the 6 helices are indicated by numbers. The arrows represent β -strands.

than in cytochrome b_5 . The heme group is also oriented somewhat differently in the two structures.

The $\beta_8\alpha_8$ motif of the flavin binding domain has been observed in a number of proteins of diverse character (40). Three of these, TMADH, flavocytochrome b_2 and glycolate oxidase (41), contain FMN. The folding topologies of the $\beta_8\alpha_8$ barrel domains of these 3 enzymes are compared in figure 13. The FMN is bound to all three enzymes in a very similar fashion. When the overall $\beta_8\alpha_8$ barrel structure of TMADH is compared with flavocytochrome b_2 , the rms deviation of about 180 equivalent C_α atoms is approximately 2.7 Å. This is slightly smaller than the deviation between C_α atoms in comparisons between functionally unrelated $\beta_8\alpha_8$ -containing structures. The close correspondence of the FMN positions in the 2 enzymes indicates

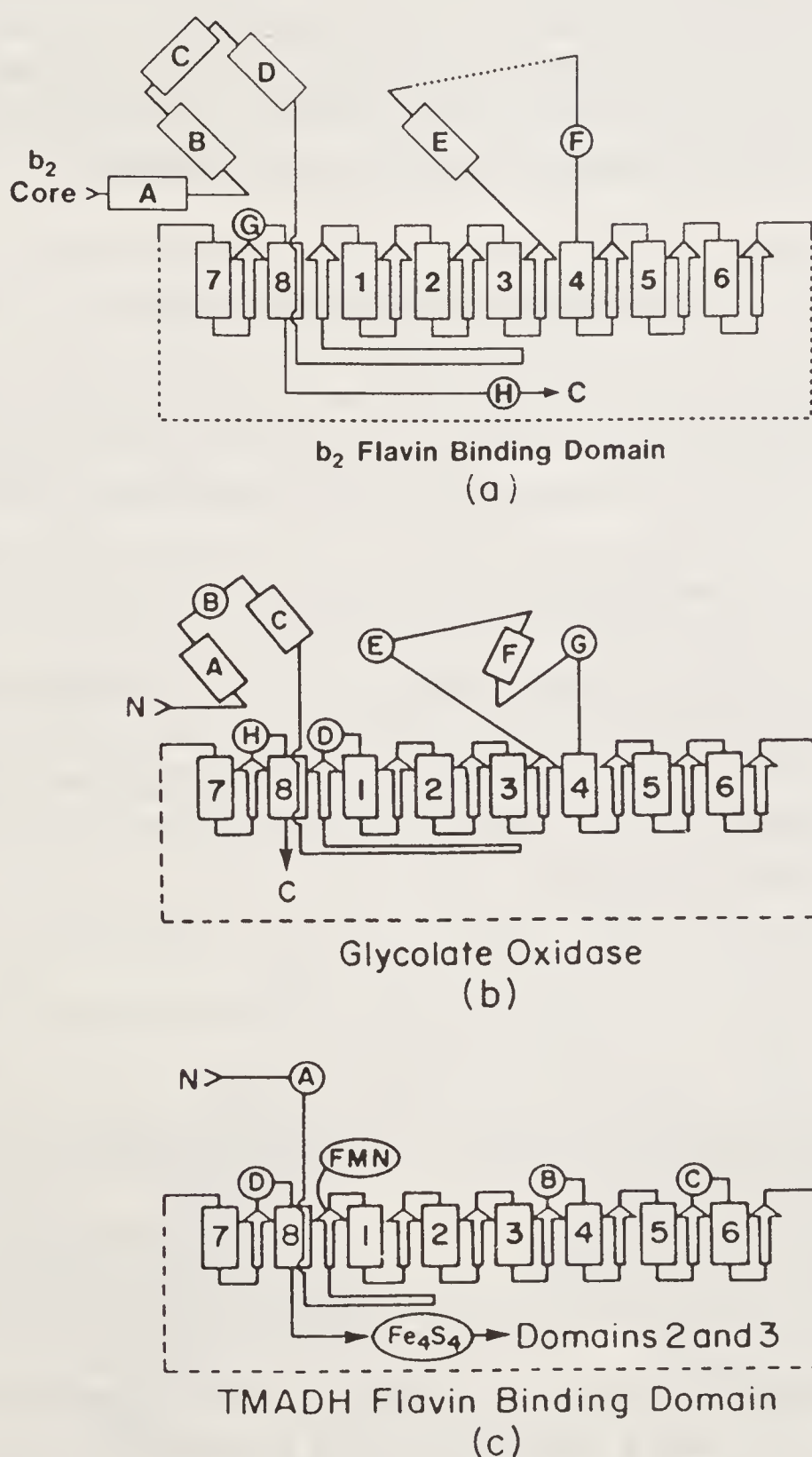


Figure 13. Folding topology of the flavin-binding domains of three FMN-containing enzymes. (a) Flavocytochrome b_2 . The dotted line between helices E and F represents a short disordered segment. (b) Glycolate oxidase (c) Trimethylamine dehydrogenase.

that the covalent binding of FMN in TMADH has little effect on its mode of binding compared to the non-covalent mode of binding.

Glycolate oxidase is much closer in structure to flavocytochrome b_2 . A detailed comparison of their 2 structures has been carried out (42). Their structural similarities extend to the loop regions between β -strands and α -helices and even outside the barrel structures. When 311 α -carbons are compared in the 2 structures, the rms deviation between equivalent positions is 0.93 Å and the sequences are 37% identical. However, when their active sites are compared, the flavin ring of FMN is found to be oriented significantly differently in the 2 structures, even though the backbone and many of the side chain positions are in very similar locations. In flavocytochrome b_2 a hydrogen bond is formed between N5 of FMN and the main chain amide of Ala 198 in β -strand 1. In glycolate oxidase the flavin ring system is tilted away from the strand, so that a pocket near the FMN ring is formed where a water molecule is bound. Model building experiments suggest that this site could accommodate a hydroperoxide intermediate during the reoxidation of glycolate oxidase by oxygen. This site would be inaccessible in flavocytochrome b_2 eliminating the oxidase pathway and favoring the electron transferase function of the enzyme.

p-Cresol Methylhydroxylase

p-Cresol methylhydroxylase (PCMH) is a flavocytochrome *c* found in the periplasmic space of certain pseudomonads. It catalyzes the first step in the oxidation of *p*-cresol to *p*-hydroxybenzaldehyde (43). During catalysis, hydroxylation of *p*-cresol occurs by the abstraction of two hydrogen atoms by the flavoprotein after which the putative *p*-quinone methide intermediate is attacked by water to yield *p*-hydroxybenzyl alcohol (44). Two electrons are then transferred one at a time to the heme on the cytochrome subunit and then to an acceptor protein *in vivo*. The acceptor protein has been identified as an azurin (45). The product of the reaction, *p*-hydroxybenzyl alcohol, can also serve as substrate for the enzyme in a dehydrogenation reaction yielding *p*-hydroxybenzaldehyde. *In vitro*, phenazine methosulfate can act as an artificial electron acceptor from the protein.

PCMH has a molecular weight of approximately 116,000 daltons. It can be resolved by isoelectric focusing into two components which differ by about 0.7 in *pI* (46). One component, of subunit M_r 49,000, contains FAD covalently bound through the 8- α position (47) to a tyrosine residue. The other, of M_r 9,000, is a *c*-type cytochrome. The amino acid sequences of the complete cytochrome subunit, the first 56 residues of the flavoprotein subunit and the FAD-binding heptapeptide are known (48).

Structure Analysis

The structure of PCMH was solved initially at 6.0 Å resolution (49) and extended to 3.0 Å resolution (50) using data recorded by area detector from crystals of the native protein and 2 heavy atom derivatives. The crystals are orthorhombic, space group $P2_12_12_1$ with cell parameters $a = 140.3$ Å, $b = 130.6$ Å and $c = 74.1$ Å. There is one complete $\alpha_2\beta_2$ tetramer in the crystallographic asymmetric unit.

Heavy atom positions for both derivatives were obtained by automatic analysis

(51) of the difference Patterson function. These positions were used to calculate the X-ray phases by the MIR method. A limited amount of anomalous scattering data was also included in the phase determination process. The collection of heavy atom positions was used to define the position of the non-crystallographic 2-fold axis. The X-ray phases were improved by a solvent leveling procedure (15) and the final map was averaged about the molecular 2-fold axis (14). The electron density was plotted as a minimap. The paths of the polypeptide chains of the cytochrome and of the flavoprotein subunits were traced. The positions of 74 alpha carbons were identified for the cytochrome subunit and 462 for the flavoprotein subunit. Electron density for the heme group and the FAD were also clearly visible in the minimap. The alpha carbon coordinates were transferred to a molecular graphics system and used to fit the amino acid sequence of all 78 amino acids of the cytochrome subunit. A polyalanine model for the flavoprotein subunit was also constructed. The flavin-binding heptapeptide was fitted to the electron density, starting from the position of the tyrosine residue. Both directions along the polypeptide chain were tested with the heptapeptide model and one of these gave a clearly superior fit to the density. This procedure established the amino terminal to carboxyl terminal direction for the polypeptide chain.

The electron density map has been improved (52) by including a third derivative and by remeasuring one of the earlier derivatives, with care being taken to record full anomalous scattering data. The final figure of merit for the protein phases improved from 0.55 to 0.62. The phases were further refined by cyclic averaging (14). The earlier model (50) was generally verified and some marked improvements in the stereochemistry, particularly in the cytochrome domain, were made.

Structure of the Molecule

The $\alpha_2\beta_2$ tetramer of PCMH is shown in figure 14. The molecule is a prolate spheroid of approximate dimensions $110 \text{ \AA} \times 75 \text{ \AA} \times 60 \text{ \AA}$. The flavoprotein subunits are tightly packed about the molecular 2-fold axis. The subunit interface is roughly rectangular with an area of approximately 2000 \AA^2 . The cytochrome subunits lie on the periphery of the flavoprotein dimer and are nestled into depressions on its surface.

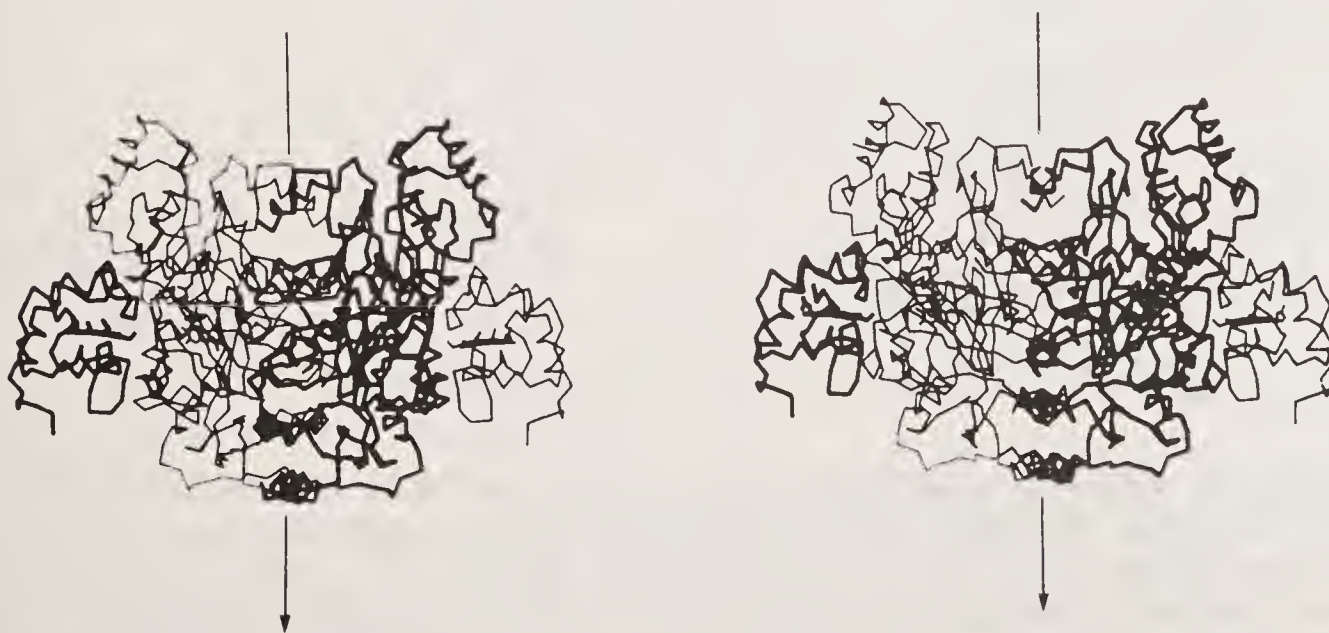


Figure 14. Stereoscopic view of the $\alpha_2\beta_2$ tetramer of PCMH. The molecular 2-fold axis is vertical.

The iron atoms in the two heme groups are separated by about 76 Å while the center to center distance between flavin rings is about 43 Å.

An α -carbon diagram of the cytochrome subunit is shown in figure 15. It consists of a compact polypeptide chain which is essentially wrapped about the heme group. It shares in common with all c-type cytochromes a number of features characteristic of the "cytochrome c fold" (53). The orientation shown in figure 15 is the standard "front view" commonly used for this family. There are 3 helical segments formed by residues 8–14, 36–45 and 62–72, and 3 reverse turns, segments 2–5, 47–50 and 55–58. The remainder of the chain is in irregular or extended conformation.

The heme group is attached covalently through its vinyl groups to Cys 15 and Cys 18 of the cytochrome subunit, respectively. Coordinate bonds are also formed between the heme iron and both His 19 and Met 50. Most of the heme group appears to be shielded from solvent, even in the isolated subunit. Exceptions are the methyl and vinyl groups at position 3 and 4, the latter covalently attached to Cys 18, and the two propionate groups.

The overall conformation and secondary structure of the flavoprotein subunit is shown in figure 16. It is folded into 3 domains. The N-terminal domain consists of approximately the first 215 residues and envelopes most of the FAD cofactor. The central domain consists of the next 208 residues and covers the flavin ring. The C-terminal domain consists of the last 39 residues and lies against the N-terminal domain, covering the adenosine portion of the FAD.

The N-terminal domain contains 4 α -helices, A–D, and the first 11 beta strands. The first 2 helices are close to the N-terminus and are 2 to 3 turns in length. The other 2 are shorter and lie between extended chains. The 11 beta strands are divided into a 4-stranded mixed β -sheet near the N-terminus (strands 1–4), a 5-stranded antiparallel beta sheet (strands 5, 6, 9–11) and a 2-stranded β -ribbon (strands 7 and 8) which interrupts the 5-stranded β -sheet. The remainder of the N-terminal domain consists of turns and extended chains. The ADP-ribose part of FAD is extended and

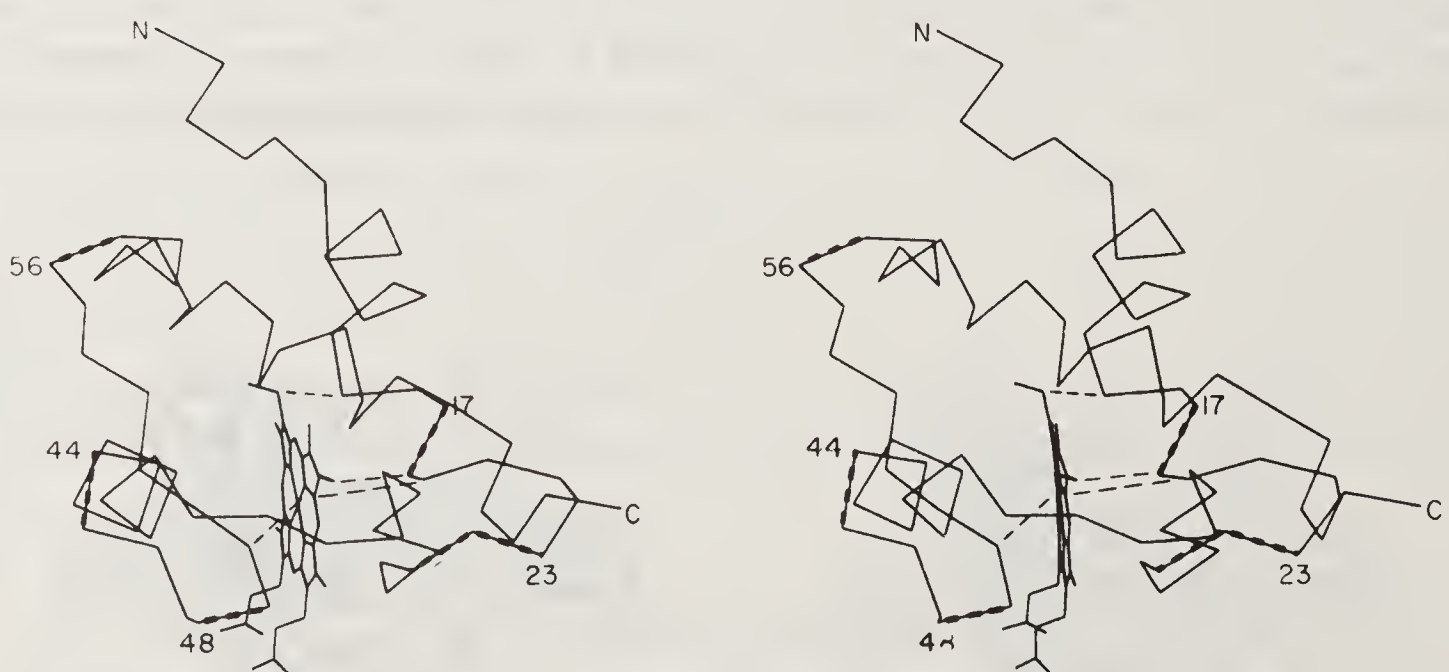


Figure 15. Backbone α -carbon tracing of the cytochrome subunit of PCMH in the "standard" orientation. The N and C-termini are indicated. Bonds from the 4 heme ligands, Cys 15, Cys 18, His 19 and Met 50 are drawn as lightly dashed lines. The 5 polypeptide segments in close contact with the flavoprotein subunit, 17–18, 23–25, 44–45, 48–49 and 56–57 are heavily dashed. The numbering of the first α -carbon of each segment is also indicated.

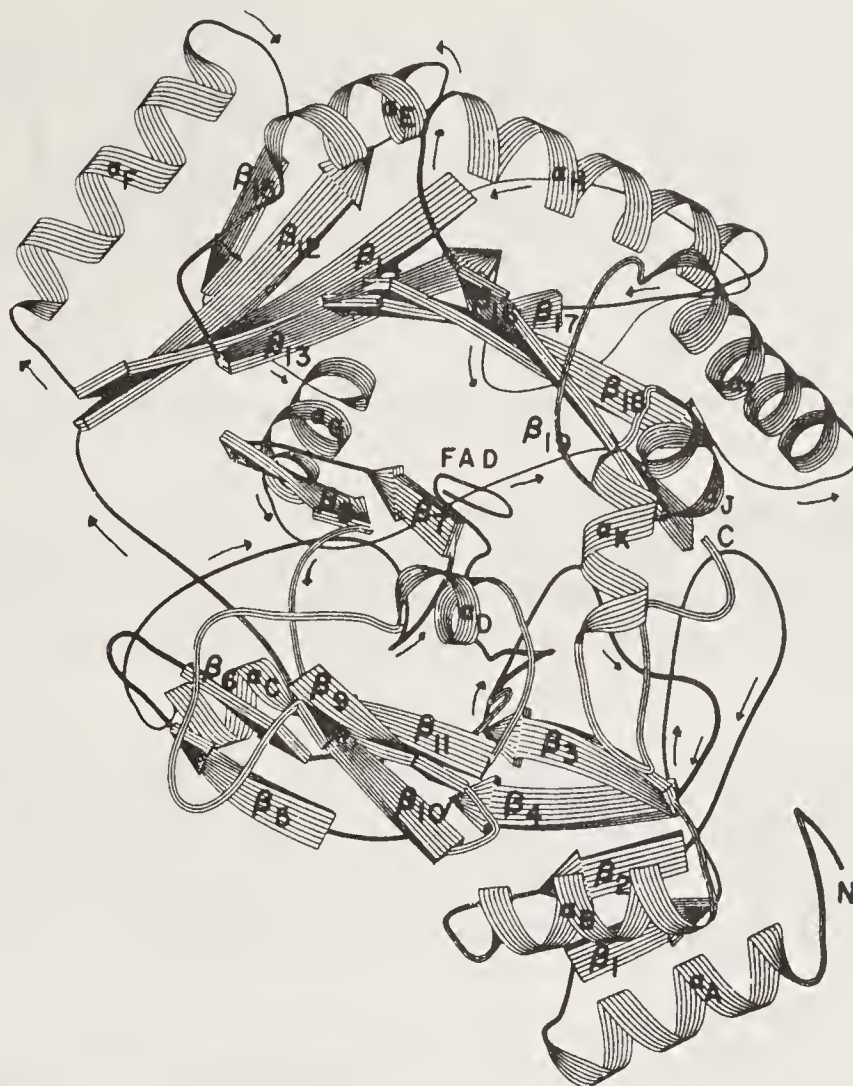


Figure 16. Cartoon of the flavoprotein subunit indicating elements of secondary structure. The positions of 11 helices (A–K) are indicated by spirals and of 18 of the 19 β -sheets by arrows. The N and C termini are also indicated. The FAD moiety is represented by a bentpaperclip-like figure. This picture was produced by the computer program RIBBON written by John Priestley.

lies in a groove in the N-terminal domain running parallel to several extended chains. This ADP-ribose-binding groove is covered by the small C-terminal domain, which contains 2 helices, J and K (figure 16). The flavin ring protrudes from the body of domain 1. Its *re* face lies against elements of domain 1 while its *si* face is open to the interior of domain 2.

The central domain is composed of an 8-stranded antiparallel β sheet, strands 12–19, which forms a dome over the isoalloxazine ring of the flavin (figure 16). The dome is covered on the outside by a side-by-side antiparallel arrangement of 4 α -helices, E, F, H and I and contains a 5th helix, G, on the inside. The site of covalent flavin attachment is at position 336 and is located about 4 residues before the beginning of strand 16.

The topology of the β -sheet of the central domain is shown schematically in figure 17. The topological order of the strands in the β -sheet is 15, 12, 14, 13, 17, 18, 16, 19 and of the helices in the side-by-side arrangement is F, E, H, I. In Richardson's topological notation (54) the connectivity is $-2x, +1, +2x, -6x, +2x, -1, -2x$. The motif of the dome consists of 2 repeating structures each containing 4 antiparallel β -strands and 2 antiparallel α -helices, connected by an additional helix. The two structures are related by a 2-fold axis perpendicular to the sheet lying between strands 13 and 17. Thus there appears to have been a structural duplication in the evolutionary development of this domain.

Although rare, the basic 4-stranded antiparallel motif found in the flavoprotein

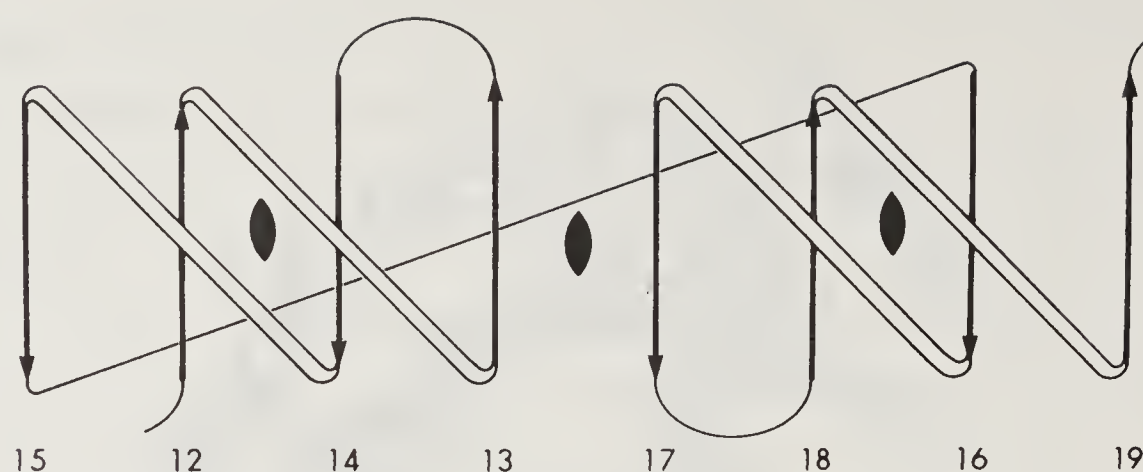


Figure 17. Topological diagram of domain 2 of PCMH viewed perpendicular to the β -sheet. The thick diagonal lines represent crossover connections above the page while the thin diagonal line indicates a crossover connection below the page. The topological 2-fold axes perpendicular to the page are indicated by the symbol ◆.

subunit is not unique. It has been observed in the 8-iron ferredoxin (55) and as part of the N-terminal domain of bacterial ribulose-1,5-bisphosphate carboxylase (56). However, the duplication to the 8-stranded form, forming the FAD-containing dome-like structure of PCMH does appear to be unique.

Cytochrome-Flavoprotein Interactions

The relative positions of the flavoprotein and cytochrome subunits are shown in figure 18. The common interface between the two subunits seems quite extensive. The thioether-containing edge of the heme group and the whole flavin group appear to be well shielded from solvent. The contributions of the cytochrome subunit to the interface of the complex are mostly neutral or hydrophobic residues. Although the cytochrome subunit is quite acidic ($pI = 4.6$) the acidic side chains lie mostly on the back and side of the molecule, some forming salt bridges to basic side chains. One

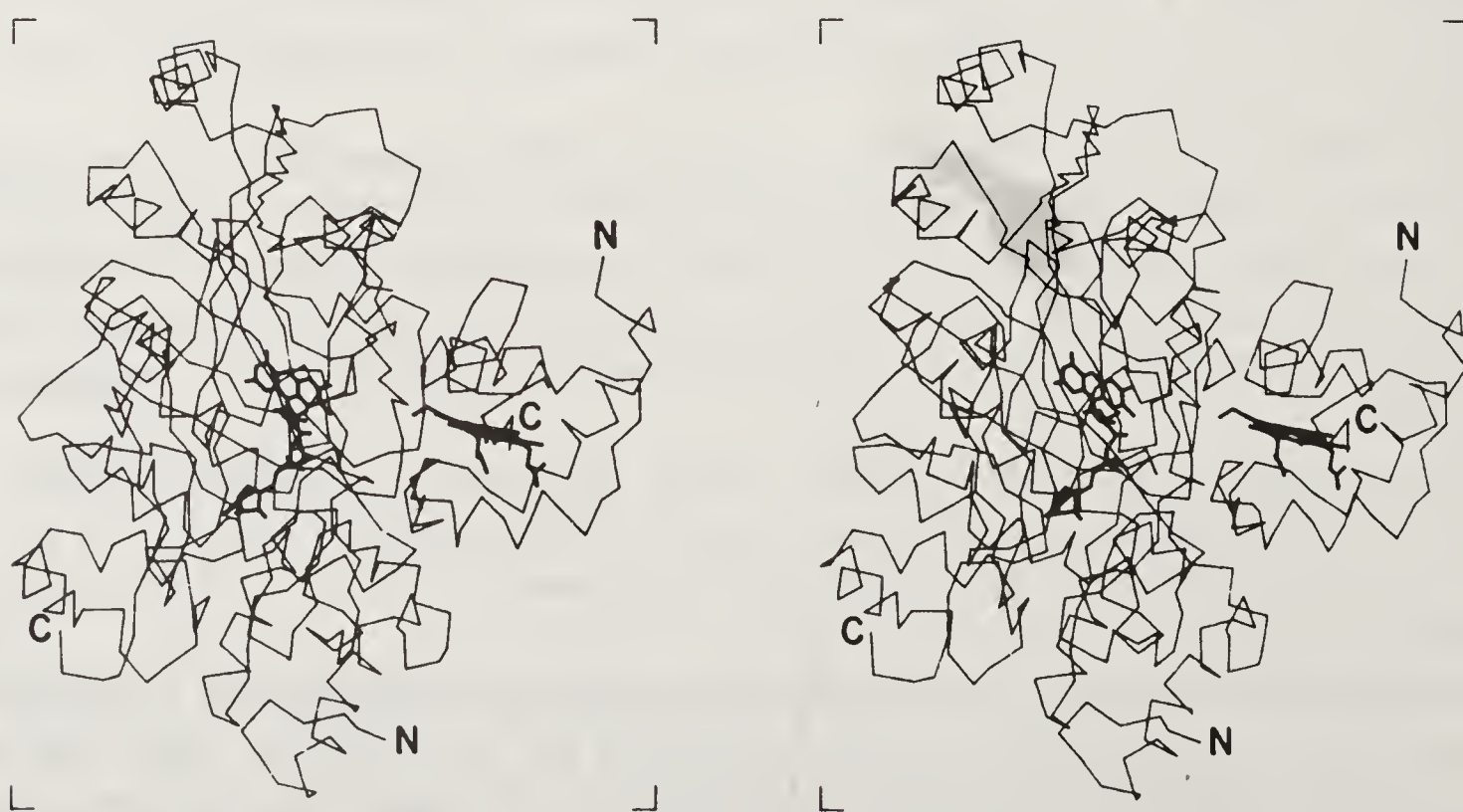


Figure 18. Stereograph showing the flavoprotein and cytochrome subunits of *p*-cresol methylhydroxylase. The polypeptide chain tracing shows the C_{α} positions. Both pairs of N- and C-termini are indicated. The heme and FAD are also shown.

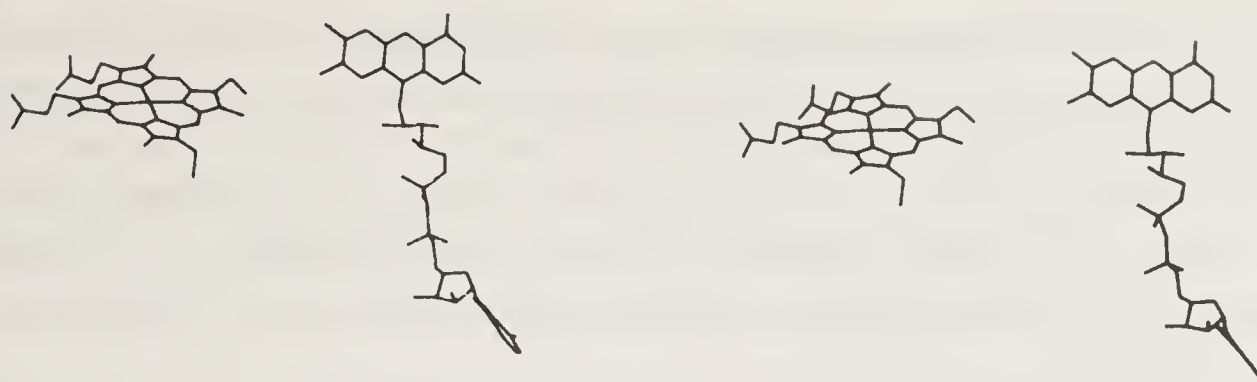


Figure 19. Stereoscopic view of the isolated heme and flavin cofactors in PCMH.

exception is Lys 13 which lies at the edge of the interface and appears to interact with a large side chain at position 124 of the flavoprotein subunit. The other is His 17 located in the center of the interface which appears to interact only with peptide backbone atoms. The region surrounding the partially exposed heme propionates is exposed to solvent in the complex. This area may serve as the binding and/or electron transfer site for the natural electron acceptor protein for PCMH.

The relative orientations of the heme and flavin cofactors are shown in figure 19. The heme iron is about 14 Å from the flavin ring and the closest approach of the two groups is about 8 Å. The benzenoid portion of the isoalloxazine ring is oriented toward the thioether-containing edge of the heme group. The flavin and heme planes are distinctly non-coplanar, making an angle of about 65° to each other.

*Binding of *p*-Cresol*

A crystal of PCMH has been soaked for 2 days in 20 mM *p*-cresol and used for data collection at 3.0 Å resolution. The largest difference density is located near the



Figure 20. Stereograph of FAD and the C $_{\alpha}$ backbone for residues 215 to 395 of the flavoprotein subunit, which forms most of domain 2. The difference density close to the flavin ring from a crystal soaked in *p*-cresol, with a model for *p*-cresol superimposed, is also shown. The numbering for the C $_{\alpha}$ backbone of the flavoprotein subunit begins at 101.

flavin ring. A molecule of *p*-cresol has been modeled into the density (figure 20). The substrate binding site is located on the *si* face of the flavin ring opposite the inner face of the antiparallel β -sheet of domain 2. It is in van der Waals contact with the flavin ring, but is inclined to it by about 30° . Each end of the *p*-cresol molecule is surrounded by 2 or 3 bulky side chains which resemble aromatic residues. Since the nature of these side chains is unknown, the orientation of *p*-cresol in the binding pocket could not be determined.

Comparison with other Proteins

The structure of the flavoprotein subunit of PCMH is unique among flavoenzymes. A general comparison of its structure with other flavoenzymes is presented in discussion and conclusions.

The cytochrome subunit of PCMH clearly belongs to the small bacterial subclass of cytochromes (57). These are acidic proteins with *pI* below 5.0 and contain about 80 or 90 amino acids. There are 2 small bacterial cytochromes whose structures have been refined to 2.5 Å resolution or better, cytochrome c_5 from *Azotobacter vinlandii* (58) and cytochrome c_{551} from *Pseudomonas aeruginosa* (59). Cytochrome c_5 contains 83 residues and has a *pI* of 4.4. Cytochrome c_{551} contains 82 residues and has a *pI* of 4.7. The PCMH cytochrome subunit is the smallest at 78 residues and has a *pI* of 4.55 to 4.70.

The distribution of charged groups among the 3 small bacterial cytochromes is also quite similar. In all three, the patch of negative charges at the rear of the molecule (as identified in figure 15) is maintained, as is the more even distribution of positive and negative charges on either side. The presence of basic groups at the bottom near the propionic acid groups and the generally neutral character of the front of the cytochrome surrounding the exposed heme face is also a common feature. This charge distribution supports the idea that the interaction of cytochrome c_5 and c_{551} with their redox partners occurs at the "front" face and is consistent with the geometry of the cytochrome c_5 dimer interaction observed in the crystalline state (58).

It is interesting to try to correlate the structural differences among the 3 cytochromes with their physiological functions. Of the five sites in PCMH which make closest contact with the flavoprotein subunit, the first 2, at 17–18 and 23–25 (see figure 16), are quite similar in structure in the 3 proteins. The next two sites, 44–45 and 48–49, are quite similar to cytochrome c_5 but differ from cytochrome c_{551} . The fifth site, 56–57, is different in all 3 proteins. This suggests that the interaction sites of cytochrome c_5 and c_{551} with their reductases occur at these corresponding positions and that the interaction of c_5 with its (unknown) redox partner is geometrically more similar to that of PCMH than is that of cytochrome c_{551} .

The major difference in structure is the large deletion in PCMH from the center of the sequence, between the two heme ligands His 19 and Met 50. In PCMH there are only 30 residues between these two ligands compared with 39 in c_5 and 43 in c_{551} . In this segment there is only 1 helix in PCMH compared with 2 helices in each of the other 2 cytochromes. The major effect of this deletion is to leave the heme propionates relatively exposed to solvent in PCMH, making this a possible interaction site for the azurin electron acceptor as discussed above.

Discussion and Conclusion

Flavoenzymes

Five distinct structural classes of flavoproteins have been observed. Two of these, the flavodoxins (60) and the $\beta_8\alpha_8$ barrel enzymes such as flavocytochrome b_2 (29) glycolate oxidase (41) and trimethylamine dehydrogenase (12), bind FMN. The other 3 classes are FAD-binding enzymes and are represented by glutathione reductase (61), by acyl Co-A dehydrogenase (62) and by PCMH, respectively. The glutathione reductase class binds the flavin in a characteristic domain similar to a Rossmann fold (19), but containing a parallel β -sheet with α -helices on one side and an antiparallel 3-stranded β -sheet on the other. In the acyl Co-A dehydrogenase the FAD is bound to a highly helical domain with the flavin ring close to a strictly β -domain. Thus, the conformation of the flavoprotein subunit of PCMH is quite different from other flavoproteins of known structure.

In all flavoproteins except acyl Co-A dehydrogenase and PCMH the phosphate or pyrophosphate group of the flavin is located at the N-terminal end of an α -helix. The positive dipole of the helix (18) is thought to stabilize the binding of the negatively charged phosphate groups. This binding feature has been found for most other dinucleotide binding proteins (63), regardless of whether positively charged side chains are present nearby. In PCMH the FAD is bound covalently, which might be thought to compensate for the lack of stabilization by the missing helix dipole. However, in acyl Co-A dehydrogenase, which also lacks this helix dipole feature (J.-J. Kim, personal communication), the flavin is bound non-covalently. The energetic need for covalent attachment of the flavin in PCMH is also questionable since the flavin binding sites in both flavocytochrome b_2 and trimethylamine dehydrogenase are nearly identical, in spite of the fact that the flavin is non-covalently attached in the former but covalently attached in the latter.

Electron Transfer Complexes

The three structures discussed contain considerable structural information on electron transfer complexes and provide a basis for evaluation of some proposed models. Although none of the structures is a true intermolecular complex between normally isolated species, they do involve interaction between different cofactors and between separate, interacting domains or subunits.

TMADH is the simplest system, containing the two cofactors, FMN and the [4Fe-4S] cluster in a single domain. The 2 cofactors are in van der Waals contact through a flavin methyl group and one of the 4 cysteine ligands to the cluster. This arrangement strongly suggests an outer sphere mechanism for electron transfer, with electrons passing directly from the 8-methyl group of FMN to the cysteine ligand. It also implicates the benzenoid ring of FMN as one of the favored pathways for electron travel in the isoalloxazine ring.

The proposed model for the flavodoxin-cytochrome c complex (1) also involved direct interaction between the solvent exposed benzenoid ring of FMN and a similarly exposed heme edge. In this case, the protein interface would exclude solvent molecules normally present in order to allow electron transfer to occur. The cofactor arrangement of TMADH supports this aspect of the model.

The structures of flavocytochrome b_2 and PCMH provide more detailed information on hypothetical complexes between electron transfer partners complexes containing macrocyclic cofactors, such as the complexes between cytochrome b_5 and cytochrome c or hemoglobin. In both flavocytochrome b_2 and PCMH, the heme and flavin cofactors are separated by about 9 Å between the edges of the ring systems. This distance is very close to the intercofactor distances proposed for the aforementioned hypothetical complexes.

Another prediction derived from the hypothetical electron transfer complexes is that the interacting protein surfaces are complementary with respect to surface topology and charge. In both flavocytochrome b_2 and PCMH, the interacting domains or subunits, respectively, are complementary with considerable contact area. In flavocytochrome b_2 there are about 30 residues in the 2 domains making hydrogen bonding and hydrophobic interactions. Furthermore, there are also a number of water molecules in the interface, several of which are directly involved in bridging the 2 domains. The PCMH structure is not yet sufficiently resolved to give detailed information on specific interaction between the subunits.

On the other hand, the structures of flavocytochrome b_2 and PCMH show that there is little ionic nature to the domain or subunit interaction. In flavocytochrome b_2 there is only one salt bridge. In PCMH, there appears to be at most one possible salt bridge between subunits, which must be verified when the amino acid sequence of the flavoprotein is completed. Other charged groups on the cytochrome subunit in the interface are oriented away from the flavoprotein subunit and would not be able to participate in salt bridges. Of course, since the domains of flavocytochrome b_2 are covalently connected, there is little need for the additional specificity provided by charge complementarity in this case. Similarly, since the intermolecular complex in PCMH is permanent, not transient, and appears to be stabilized by as yet unidentified putative hydrophobic and hydrogen bonded interactions, there appears to be no need for specific ionic interactions. Thus the small number of interdomain or intersubunit salt bridges observed in these 2 proteins does not invalidate the hypotheses of the models.

The third principle used in constructing models of electron transfer complexes is maintenance of coplanarity of cofactors. In flavocytochrome b_2 , the heme and flavin group are nearly coplanar, to within 15°. However, in PCMH, the heme and flavin are clearly not coplanar. The angle between planes is approximately 65°.

The three structures presented here also show that there is variability in the pathway for electron flow in the flavin ring. In all 3 structures, the substrate binding sites appear to favor reduction of the flavin through the N5 position, since this position is freely accessible to the binding sites. However, in 2 cases, TMADH and PCMH, flavin reoxidation appears to involve electron flow through the benzenoid ring while in the third, flavocytochrome b_2 , the reoxidation appears to involve the N5 locus.

Acknowledgement

This work has been supported by grants from the National Institutes of Health and by the National Science Foundation.

References

1. Simondsen, R. P., Weber, P. C., Salemme, F. R. & Tollin, G. (1982) *Biochemistry* **21**, 6366–6375.
2. Salemme, F. R. (1976) *J. Mol. Biol.* **102**, 563–568.
3. Poulos, T. L. & Mauk, G. R. (1983) *J. Biol. Chem.* **258**, 7369–7373.
4. Steenkamp, D. J. & Mallinson, J. (1976) *Biochim. Biophys. Acta* **429**, 705–719.
5. Lim, L. W., Mathews, F. S. & Steenkamp, D. J. (1982) *J. Mol. Biol.* **162**, 869–876.
6. Kasprzak, A. A., Papas, E. J. & Steenkamp, D. J. (1983) *Biochem. J.* **211**, 535–541.
7. Steenkamp, D. J., McIntire, W. & Kenney, W. C. (1978) *J. Biol. Chem.* **253**, 2818–2824.
8. Steenkamp, D. J. & Gallup, M. (1978) *J. Biol. Chem.* **253**, 4086–4089.
9. Steenkamp, D. J., Beinert, H. & Singer, T. P. (1978) *Biochem. J.* **169**, 361–369.
10. Steenkamp, D. J. & Beinert, H. (1982) *Biochem. J.* **207**, 233–239.
11. Stevenson, R. C., Dunham, W. R., Sands, R. H., Singer, T. P. & Beinert, H. (1986) *Biochim. Biophys. Acta* **869**, 81–88.
12. Lim, L. W., Shamala, N., Mathews, F. S., Steenkamp, D. J., Hamlin, R., & Xuong, Ng.H. (1986) *J. Biol. Chem.* **261**, 15140–15146.
13. Howard, A. J., Nielsen, C. & Xuong, Ng-H. (1985) *Methods Enzymol.* **114**, 452–472.
14. Bricogne, G. (1976) *Acta Crystallogr.* **A32**, 832–847.
15. Wang, B. C. (1982) *Methods Enzymol.* **115**, 90–112.
16. Mathews, F. S. & Lim, L. W. (1987) in *Flavins and Flavoproteins*, ed. McCormic, D. B. & Edmondson, D. E. (Walter de Gruyter, Berlin) pp. 663–672.
17. Hendrickson, W. A. (1985) *Methods Enzymol.* **115**, 252–270.
18. Hol, W. G. J., van Duijnen, P. T. & Berendson, H. J. C. (1978) *Nature (London)* **273**, 443–446.
19. Holbrook, J. J., Liljas, A., Steindel, S. J. & Rossmann, M. G. (1975) in *The Enzymes*, 3rd edition, ed. Boyer, P. D. (Academic Press, New York) Vol. 11, pp. 191–292.
20. Thieme, R., Pai, E. F., Schirmer, R. H. & Schulz, G. E. (1981) *J. Mol. Biol.* **152**, 763–782.
21. Glehn, M., Stensland, B. & Gärtner, B. (1977) *Acta Crystallogr* **B33**, 2388–2391.
22. Bellamy, H. D., Lim, L. W., Mathews, F. S. & Dunham, W. R. (1989) *J. Biol. Chem.* **264**, 11887–11892.
23. Appleby, C. A. & Morton, R. K. (1954) *Nature (London)* **173**, 749–752.
24. Capeillère-Blandin, C., Bray, R. C., Iwatsubo, M. & Labeyrie, F. (1975) *Eur. J. Biochem.* **54**, 549–566.
25. Lederer, F., Cortial, S., Becam, A. M., Haumont, P. Y. & Perez, L. (1985) *Eur. J. Biochem.* **139**, 59–65.
26. Guiard, B., Groudinsky, O. & Lederer, F. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2539–2543.
27. Ghrir, R. & Lederer, F. (1981) *Eur. J. Biochem.* **120**, 279–287.
28. Jacq, C. & Lederer, F. (1974) *Eur. J. Biochem.* **41**, 311–320.
29. Xia, Z.-X., Shamala, N., Bethge, P. H., Lim, L. W., Bellamy, H. D., Xuong, Ng.-H., Lederer, F. & Mathews, F. S. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2629–2633.
30. Xia, Z.-X. & Mathews, F. S. (1990) *J. Mol. Biol.* **212**, 837–863.
31. Phizackerley, R. P., Cork, C. W. & Merritt, E. A. (1986) *Nucl. Instrum. Methods* **A246**, 579–595.
32. Urban, P. & Lederer, F. (1984) *Eur. J. Biochem.* **144**, 345–351.
33. Lederer, F. & Mathews, F. S. (1987) in *Flavins and Flavoproteins*, eds. Edmondson, D. E. & McCormick, D. B. (Walter de Gruyter, Berlin) pp. 133–142.
34. Ghisla, S. (1982) in *Flavins and Flavoproteins*, eds. Massey, V. & Williams, C. H. (Elsevier, Amsterdam) pp. 133–140.
35. Lederer, F. (1978) *Eur. J. Biochem.* **88**, 425–431.
36. Massey, V., Muller, F., Feldberg, R., Schuman, M., Sullivan, P. A., Howell, L. G., Mayhew, S. G., Matthews, R. G. & Foust, G. P. (1969) *J. Biol. Chem.* **244**, 3999–4006.
37. Mozzarelli, A., Tegoni, M. & Rossi, G. L. (1984) in *Flavins and Flavoproteins*, eds. Bray, R. C., Engel, P. C. & Mayhew, S. G. (Walter de Gruyter, Berlin, New York) p. 251.
38. Muller, F. & Massey, V. (1969) *J. Biol. Chem.* **244**, 4007–4016.
39. Mathews, F. S., Argos, P. & Levine, M. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 387–395.

40. Farber, G. K. & Petsko, G. A. (1990) *Trends Biochem. Sci.* **15**, 228–234.
41. Lindqvist, Y. & Branden, C.-I. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6855–6859.
42. Lindqvist, Y., Branden, C.-I., Mathews, F. S. & Lederer, F. (1991) *J. Biol. Chem.* **266**, 3198–3207.
43. Hopper, D. J. & Taylor, D. G. (1977) *Biochem. J.* **167**, 155–162.
44. Hopper, D. J. (1978) *Biochem. J.* **175**, 345–347.
45. McIntire, W., Hopper, D. J. & Singer, T. P. (1985) *Biochem. J.* **228**, 325–335.
46. McIntire, W. S., Koerber, S. C., Bohmont, C. W. & Singer, T. P. (1984) in *Flavins and Flavoproteins*, eds. Bray, R. C., Engel, P. C. & Mayhew, S. G. (DeGruyter and Co., N. Y.) pp. 643–654.
47. McIntire, W. S., Edmondson, D. E., Hopper, D. J. & Singer, T. P. (1981) *Biochemistry* **20**, 3068–3075.
48. McIntire, W., Singer, T. P., Smith, A. J. & Mathews, F. S. (1986) *Biochemistry* **25**, 5975–5981.
49. Shamala, N., Lim, L. W., Mathews, F. S., McIntire, W., Singer, T. P. & Hopper, D. J. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4626–4630.
50. Mathews, F. S., Chen, Z.-W., Bellamy, H. D. & McIntire, W. S. (1991) *Biochemistry*, **31**, 238–247.
51. Terwilliger, T. C., Kim, S. H. & Eisenberg, D. (1987) *Acta Crystallogr.* **A43**, 1–5.
52. Mathews, F. S., Chen, Z.-W., Bellamy, H. D. & McIntire, W. S. (1991) in *Flavins and Flavoproteins*, ed. Curti, B. (DeGruyter and Co., N. Y.) pp 805–808.
53. Almassy, R. J., & Dickerson, R. E. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2674–2678.
54. Richardson, J. S. (1977) *Nature (London)* **268**, 495–500.
55. Adman, E. T., Sieker, L. C. & Jensen, L. H. (1973) *J. Biol. Chem.* **248**, 3987–3996.
56. Schneider, G., Lindqvist, Y., Branden, C.-I. & Lorimer, G. (1986) *EMBO J.* **5**, 3409–3415.
57. Dickerson, R. E. (1980) *Sci. Am.* **242**, 98–110.
58. Carter, D. C., Melis, K. A., O'Donnell, S. E., Burgess, B. A., Furey, W. F., Jr., Wang, B. C. & Stout, C. D. (1985) *J. Mol. Biol.* **184**, 279–295.
59. Matsuura, Y., Takano, T. & Dickerson, R. E. (1982) *J. Mol. Biol.* **156**, 389–409.
60. Ludwig, M. L., Pattridge, K. A., Smith, W. W., Jensen, L. H. & Watenpaugh K. (1982) in *Flavins and Flavoproteins*, eds. Massey, V. & Williams, C. H., Jr. (Elsevier, North Holland, New York) pp. 19–27.
61. Karplus, P. A. & Schulz, G. E. (1987) *J. Mol. Biol.* **195**, 701–729.
62. Kim, J.-J. P. & Wu, J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6677–6681.
63. Wierenga, R. K., De Maeyer, M. C. & Hol, W. G. J. (1985) *Biochemistry* **24**, 1346–1357.

Crystallographic Investigations of Glycolytic Enzymes from *Trypanosoma brucei*: Potential Starting Points for the Design of New Sleeping Sickness Drugs

Wim G. J. Hol¹, Fred M. D. Vellieux¹, Christophe L. M. J. Verlinde¹,
Rik K. Wierenga², Martin E. M. Noble² and Randy J. Read³

¹BIOSON Research Institute, Department of Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands

²EMBL, Postfach 10.2209 Meyerhofstrasse 1, 6900 Heidelberg, FRG

³Department of Medical Microbiology and Infectious Diseases, University of Alberta, Edmonton, Alberta, Canada T6G 2H7

Introduction

Living organisms owe their unique characteristics to the properties of a large number of complex biomacromolecules. Among these, proteins occupy a special position because they carry out an incredible range of different and often crucial functions. Their importance is shown, for instance, by the fact that several diseases can be traced back to minute alterations in a single protein molecule. An example is the point mutation in sickle cell hemoglobin which leads to sickle cell anemia (1).

In addition to inherited defects, errors in the amino acid sequence of a protein can be caused by damage at the DNA level which occurs during the life of a human being. An example is the point mutation of the *ras* gene, leading to mutated p21 protein, which is closely associated with some forms of cancer (2). Human disease is thus often caused by errors in proteins themselves, either inherited or acquired. If drugs are to be useful in such cases they should counteract the effects of erroneous protein structures.

Many other diseases are, of course, not due to deficiencies of human proteins, but to invasions by harmful organisms. In such cases drugs have to interfere with crucial pathways of the pathogenic organism without, ideally, affecting the human host. Sufficient selectivity of the compounds used is the goal to be achieved (3).

This goal has been reached in many cases by approaches which consisted mainly of trial-and-error procedures no matter how sophisticated the methods employed were. The advent of a number of new techniques – ranging from recombinant DNA technology, rapid gene sequencing techniques, modern protein purification and protein microsequencing methods, oligonucleotide synthesis, X-ray crystallography, computer graphics, molecular modelling to molecular dynamics calculations – are now making it possible to try to design new drugs based upon three-dimensional structures of target proteins elucidated at the atomic level (see, e.g., ref. 4).

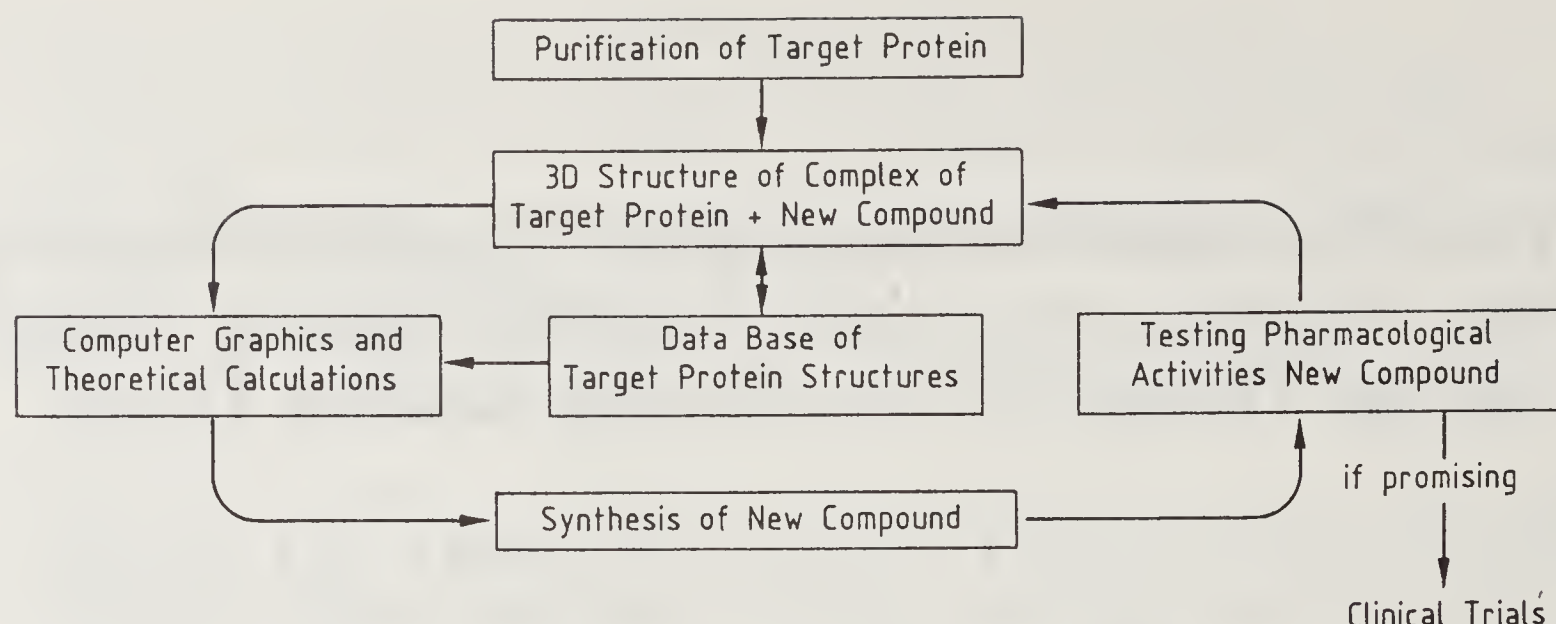


Figure 1. Schematic drawing of the "rational drug design cycle" (4).

Of course, such a rational drug design approach has still several trial and-error components because of:

- (i) Our incomplete understanding of intermolecular interactions which may e.g. result in unexpected modes of binding of newly synthesized compounds to the target protein;
- (ii) The potential drug may not reach the target protein *in vivo*;
- (iii) Metabolic processes may transform the potential drug into an ineffective compound;
- (iv) The newly synthesized compound may be toxic, teratogenic, mutagenic or carcinogenic. Therefore, it is useful to consider a rational drug design *cycle* (figure 1) which leads to a gradual increase of understanding of the mode of binding of a range of compounds to the potential target protein. Once such a database of three-dimensional structures of potential drugs bound to target proteins, plus associated binding constants, is obtained, the efficiency of the drug design procedure should be significantly increased.

In this contribution we describe the current status of a collaborative project aimed at development of therapeutically useful compounds for the treatment of African trypanosomiasis, or sleeping sickness, on the basis of three-dimensional structures of key parasite enzymes.

Sleeping Sickness or African Trypanosomiasis

Sleeping sickness is a serious threat to the health of the people living in large parts of Africa. If untreated it is virtually always fatal. Treatment requires hospitalization, because the drugs currently in use are very toxic (5). About 50 million human beings in 34 African countries are at risk of developing the disease (6), which is caused by a unicellular protozoan haemoflagellate, the trypanosome. Two subspecies of *Trypanosoma brucei* are the causative agents: *T. b. gambiense* and *T. b. rhodesiense*. Trypanosomes also cause an infectious cattle disease ("nagana"). Of greatest importance in this respect are *T. vivax* and *T. congolense*. The various trypanosomal species belong to the family Trypanosomatidae, which is a subdivision of the Kinetoplastida.

As trypanosomes affect both the physical as well as the economical well-being of the inhabitants of tropical Africa, the World Health Organization (WHO) has included this disease in its special Program for Research and Training in Tropical Diseases (6).

Trypanosomes are transmitted by a limited number of *Glossina* species, insects commonly known as the tsetse fly. After a bite by an infected fly, the parasites move to the bloodstream of the host. *T. brucei* cells, quite remarkably, do not hide in any of the host blood cells. They are in this respect quite different from e.g. the malaria parasite which is found in the erythrocyte. The trypanosomes multiply in the blood by binary fission. Eventually, other organs become infected whereby, surprisingly, the organism is able to cross the blood-brain barrier. The invasion of the central nervous system causes the final symptoms of sleeping sickness: insomnia, epileptic attacks, coma and eventually death (5).

In view of the shortcomings of the available drugs (7), we have embarked on a project which aims at obtaining new drugs by exploiting peculiarities of the trypanosomal biochemical machinery.

Trypanosomes

Trypanosomes have a quite complicated lifecycle, both in the tsetse fly and in the host (5). Major morphological changes occur during this cycle. In this chapter we will focuss on the bloodstream form which is schematically depicted in figure 2. This form has several unusual features which have recently been reviewed by Oppenheimer (8) and some of which will be briefly discussed in the next few sections.

Kinetoplast DNA

This is a network of small and large circular DNA-molecules in mitochondria. Each mitochondrion has one such network, which represents between 5 and 30% of total cellular DNA (8). The “mini-circles”, $\sim 0.3 \mu\text{m}$ in length, make up 90–95% of the network mass. The “maxi-circles”, $\sim 16 \mu\text{m}$ in length, code for mitochondrial proteins. Some species, such as *T. brucei evansi* and *T. brucei equiperdum* have lost their maxi-circles, therefore they cannot develop a functional mitochondrial respiratory chain. Consequently they have lost the ability to develop in the insect host.

Exciting recent discoveries have led to deeper insight as to the function of these circular DNA molecules. Some mitochondrial proteins of trypanosomes appear to be encoded on the DNA in a most unusual fashion. For example, the gene for a cytochrome c oxidase subunit is missing numerous thymines – up to 50% (!) – when compared with the eventual amino acid sequence (9–11). The gene is therefore called a ‘cryptogene’. Extensive RNA editing occurs in the trypanosome which appears to involve the use of numerous short stretches of ‘guide RNA’s’, encoded elsewhere on the DNA (12, 13). In a fascinating editing process the missing bases are somehow inserted, and a few deleted, so that a proper mature RNA eventually arises.

Clearly, this process is distinctly different from that of mammalian protein biosynthesis. Since the mitochondrion is poorly developed in the bloodstream form of the parasite (8), RNA-editing may occur for only a small number of proteins in the bloodstream form of the parasite. However, if such proteins would turn out to

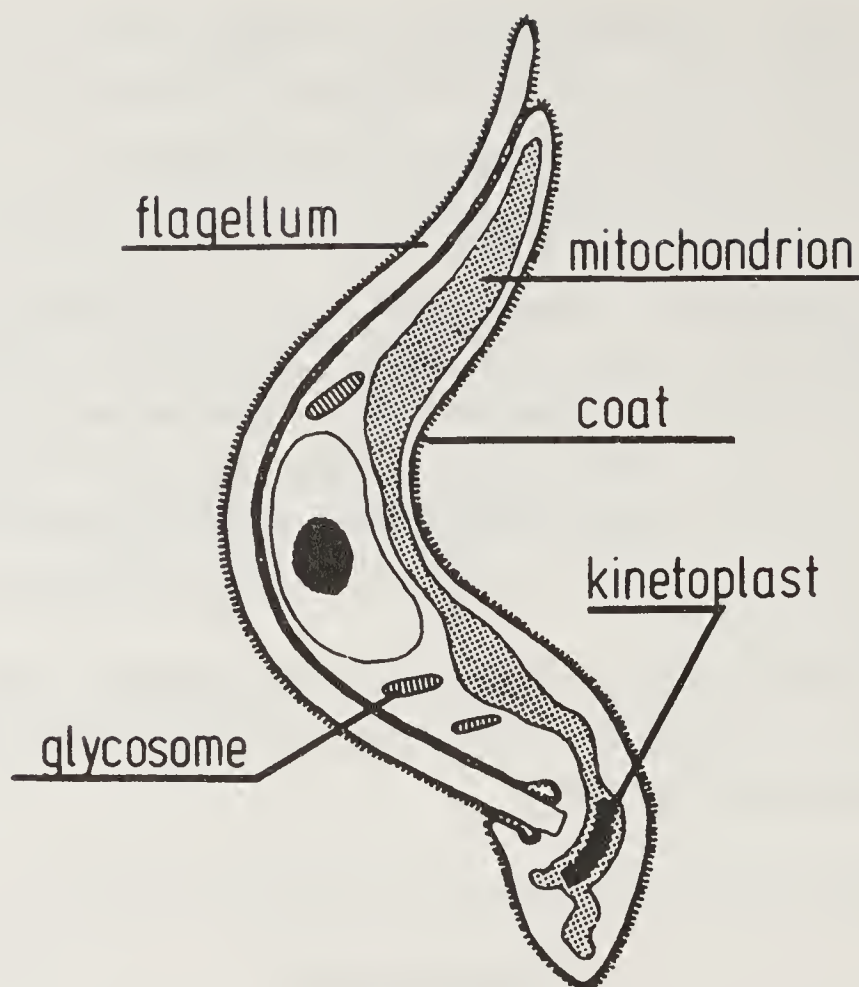
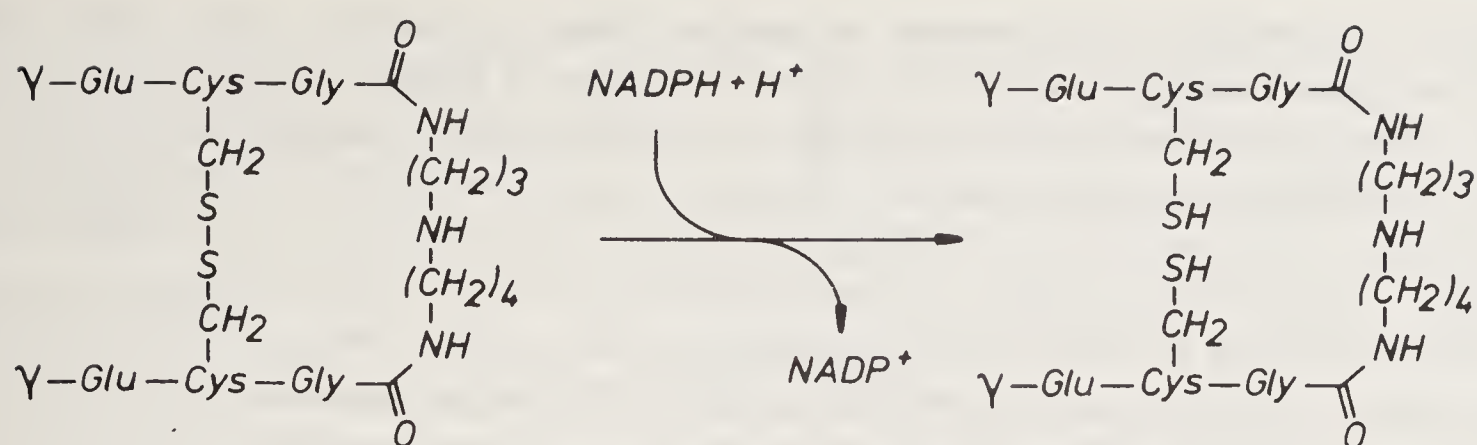


Figure 2. Schematic drawing of the bloodstream form of the trypanosome which is approximately $10\text{ }\mu\text{m}$ long. Due to the flagellum, trypanosomes in the blood move continuously and rapidly. Three peculiar features of trypanosomes – the surface coat, the kinetoplast and the glycosome – are discussed more extensively in the text. The black dot represents nuclear DNA. The grey organelle running along the entire length of the trypanosome is the mitochondrion, which is hardly developed in the bloodstream form of *Trypanosoma brucei*. Drawing kindly provided by Dr. F. R. Opperdoes.

be essential for the parasite's metabolism these unique large scale editing processes would be intriguing targets for drug design. So far, the proteins and nucleoprotein complexes involved have not yet been characterized at the atomic level and, hence, will not receive further attention in this overview.

Trypanothione and Polyamine Metabolism

In trypanosomes, the ratio of intracellular levels of reduced and oxidized thiols is maintained by a 'novel' enzyme, trypanothione reductase (14, 15). Trypanothione is a glutathione–spermidine conjugate (figure 3), whose structure has been elucidated by Fairlamb *et al.* (14). Trypanothione has been found in several parasitic Kinetoplastida, such as *Trypanosoma brucei*, *Trypanosoma cruzi*, *Crithidia fasciculata* and *Leishmania mexicana*. Trypanothione reductase of both *T. cruzi* (16) and *Crithidia fasciculata* (15) has been purified. It is a flavoenzyme which is homologous to glutathione reductase. The molecular weight of one subunit of trypanothione reductase, a dimeric protein, is approximately 50,000 daltons. It catalyzes the NADPH-dependent reduction of oxidized trypanothione (figure 3). The active site peptide containing the catalytically important cysteine residues is identical to the analogous peptide of human glutathione reductase. Inhibition of this unique trypanosomal enzyme will affect the



Oxidized Trypanothione

Reduced Trypanothione

Figure 3. Reaction catalyzed by trypanothione reductase. The oxidized form of trypanothione contains a 24-atom ring, which is formed by spermidine and by atoms of the glutathione-moieties.

ratio of the intracellular concentrations of reduced and oxidized thiols, which is expected to be deleterious to the trypanosome (14), because it will increase the sensitivity to reactive oxygen-containing molecules.

Recent developments are that this important trypanosomal protein from *Crithidia fasciculata*, an insect trypanosome without bloodstream form, has been overexpressed and crystallized. The structure determination of this enzyme is now underway (W. Hunter *et al.*, personal communication; J. Kuriyan, personal communication). Once high-resolution structures are available this enzyme will provide an exciting starting point for drug design, after molecular modelling procedures have created an approximate model of the true trypanothione reductase of *Trypanosoma brucei*.

The spermidine moiety of trypanothione is a polyamine. It is therefore interesting that a compound which interferes with the polyamine metabolism turns out to be a new drug against sleeping sickness. This compound is DFMO, difluoromethyl ornithine, a suicidal inhibitor of ornithine decarboxylase (17). This enzyme decarboxylates ornithine into putrescine ($\text{H}_2\text{N}-(\text{CH}_2)_4-\text{NH}_2$). Putrescine can subsequently be converted (18) by another enzyme into spermidine ($\text{H}_2\text{N}-(\text{CH}_2)_3-\text{NH}-(\text{CH}_2)_4-\text{NH}_2$) which is an essential component of trypanothione. DFMO is now in use (19) as a chemotherapeutic agent against sleeping sickness. Some disadvantages associated with the general use of this compound as a therapeutic agent are the large quantities needed to be effective, the occurrence of resistance and the relative insensitivity of *T. brucei rhodesiense* the most virulent form of the parasite for DFMO (20).

Variable Surface Glycoproteins

The bloodstream form of the trypanosome is continuously exposed to the immune system of the host. The surface of these trypanosomes is a dense coat of protein molecules that is 120–150 Å thick. It is a monolayer of identical copies of a particular coat protein, the Variable Surface Glycoprotein (VSG), which has antigenic properties (21, 22). The immune system successfully produces antibodies against this coat protein, but nevertheless the trypanosomal infection persists. This is due to the phenomenon of antigenic variation: regularly a new VSG is synthesized which eventually completely

replaces the previous coat protein. Altogether there are perhaps more than 1000 different genes coding for VSG's. Only one VSG-gene is being expressed at the same time, although at the time of switching two different VSG's distributed uniformly over the cell surface, can be detected by monoclonal antibodies (23, 24). The molecular mechanism of the expression of VSG-genes has been described in a recent review (25).

VSG's consist of two domains. Upon treatment by proteases the N-terminal domain is cleaved off and a soluble VSG-fragment can be purified. These VSG fragments form dimers with subunits of 40,000 daltons each. N-terminal domains of two different VSG's have been crystallized. Despite lack of sequence homology these domains have very similar structures as can be deduced from low resolution electron density maps (26, 27). A prominent feature in these structures are 2 long α -helices (80 Å) per subunit. Disruption of the folding of VSG's in such a manner that more conserved regions become exposed to circulating antibodies may be a potential, but rather speculative, approach to developing new trypanosomal drugs. Other potential targets are interference with the machinery which ensures copying of a new VSG gene to the expression site, and blocking the attachment of VSG's to the membrane. So far, however, the proteins involved have not been characterized in detail.

An important consequence of the effective manner by which the trypanosome avoids the immune system is that development of a protective vaccine is very difficult indeed. This is a major reason to try to develop new drugs on the basis of a vital metabolic pathway described in the next section.

Glycolysis and Glycosomes

Glucose uptake and utilization by the bloodstream form of the trypanosomes is very fast. Per mg cellular protein the glucose consumption is $\sim 85 \text{ nmol} \cdot \text{min}^{-1} (\text{mg protein})^{-1}$. This is 50 times faster than the glucose consumption of mammalian cells (28). From this turnover rate, and assuming a protein content of 10% (w/w) (29), it can be estimated that the amount of glucose consumed by trypanosomes in a period of 7 hours, is roughly equal to the total cellular weight.

In the bloodstream form of the trypanosome the single mitochondrion has been reduced to a peripheral canal without a functional Krebs' cycle. Moreover, the trypanosomes lack storage forms of metabolic energy, such as carbohydrates or "high energy phosphate"-molecules, which further increases their vulnerability for compounds interfering with glycolysis (30). Not surprisingly therefore a mixture of salicyl-hydroxamic acid and glycerol, which is expected to block glycolysis (31), kills the trypanosomes *in vitro* (32) as well as *in vivo* (33). For pharmacological reasons, such as toxicity for the host, the combination of these two compounds is obviously not an acceptable drug.

Trypanosomes have a specialized organelle, called "glycosome" by the discoverers Opperdoes and Borst (33), involved in efficient glycolysis (34). Glycosomes are microbodies with a diameter of $\sim 0.3 \mu\text{m}$ which are surrounded by a single membrane. The lipids occurring in this membrane are the common phospholipids phosphatidylcholine and phosphatidyl ethanolamine (35). The major proteins sequestered in the glycosomes are nine enzymes involved in glucose and glycerol metabolism, catalyzing the metabolic path shown in figure 4. Other glycosomal enzymes are involved in carbon dioxide fixation (36), as well as in ether lipid and pyrimidine synthesis (34). This indicates common features, and hence a possible common origin,

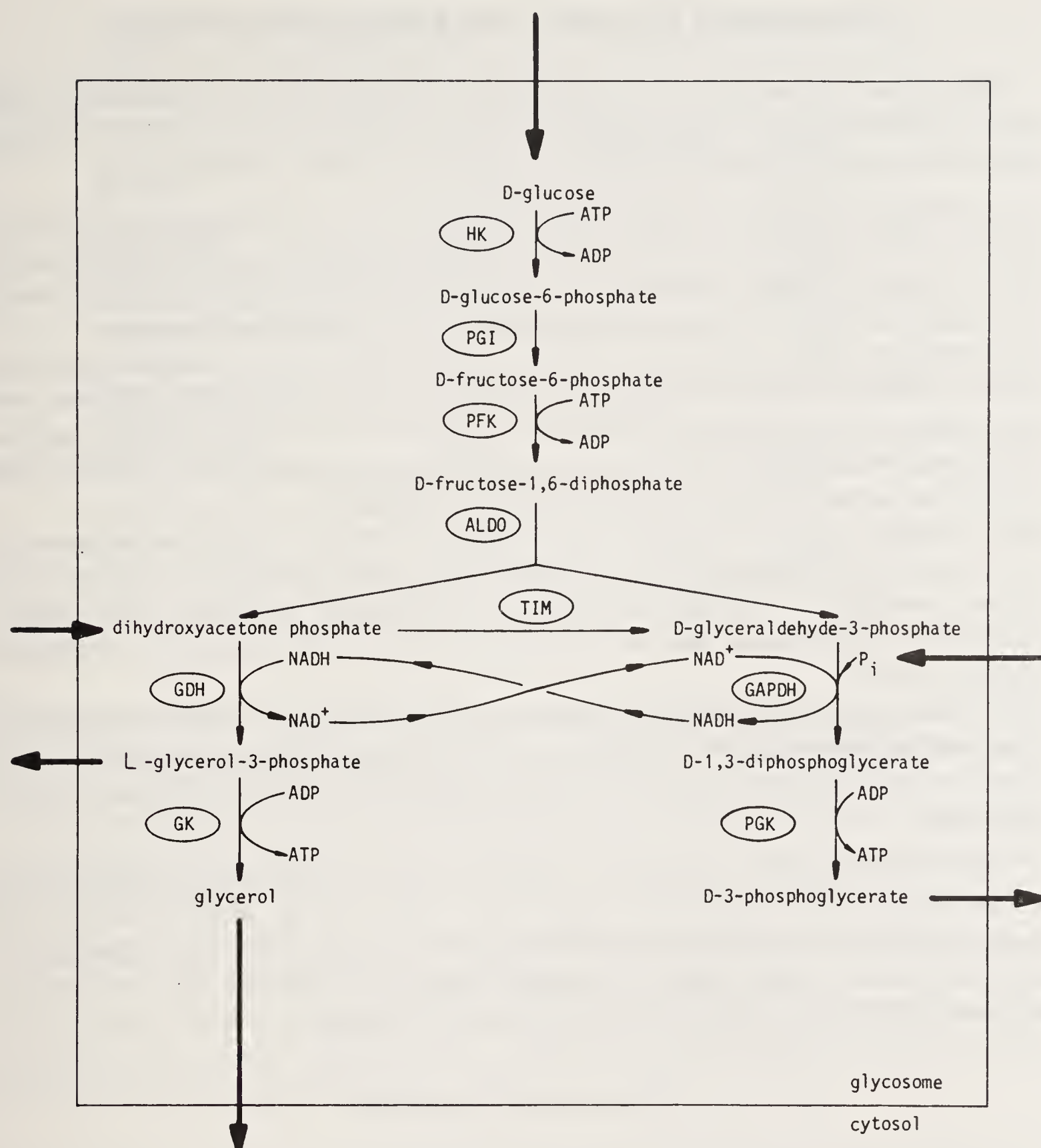


Figure 4. The glucose and glycerol metabolism in the glycosome (36). The abbreviations (table 1) of the nine enzymes are encircled. Six compounds can be exchanged between the glycosomal matrix and the cytosol (indicated by arrows): D-glucose, dihydroxyacetone phosphate, glycerolphosphate, glycerol, inorganic phosphate (P_i), and D-3-phosphoglycerate. Under aerobic conditions 1 molecule of D-glucose is converted into 2 molecules of D-3-phosphoglycerate. The net ATP-production occurs in the cytosol. Figure kindly provided by Dr F. R. Opperdoes.

with other organelles such as peroxisomes. It should be pointed out that the glycosomes have not only been found in *T. brucei*, but have also been identified in several other members of the order Kinetoplastidae, such as *T. cruzi*, *Crithidia* spp. and in four representatives of *Leishmania* spp. (see e.g. ref. 36, p. 199).

Given the importance of glycosomes for trypanosomes, it was decided to focus on glycosomal enzymes as potential targets for drug design.

Trypanosomal Glycolysis and Rational Drug Design

The unique features and crucial importance of glycolysis in trypanosomes offer several opportunities for designing new therapeutic compounds following the 'rational drug design cycle' (4) depicted in figure 1. Various possible approaches are:

- (i) Exploiting the fact that glycosomal enzymes have to be translocated into the glycosome. Interference with this membrane passage process into an organelle which has no equivalent in the human host is a potentially promising approach, although our understanding of the import mechanism is still quite limited at present;
- (ii) Exploiting differences in the catalytic centres of glycosomal enzymes and their human analogues. With knowledge of kinetic properties, amino acid sequences and three-dimensional structures it might be feasible to arrive in a rational way at compounds which bind well to glycosomal enzymes and inhibit their activity while leaving the human analogous enzymes largely unaffected;
- (iii) Interfering with the assembly of glycosomal enzymes. All of these enzymes are multimers, except phosphoglycerate kinase. For several, if not all, of these proteins proper assembly is required for activity. Compounds interfering with correct association of glycosomal, and not of human, multi-subunit enzymes are potentially useful as drugs;
- (iv) The monomeric enzyme phosphoglycerate kinase (PGK) has to make a considerable hinge bending motion during its catalytic cycle. Compounds which inhibit this motion in trypanosomal PGK may be useful as drugs as long as interference with human PGK is limited.

With these strategies in mind we will now look at glycosomal enzymes from *Trypanosoma brucei*. Ideally we would like to have at our disposal amino acid sequences, full kinetic data and three-dimensional X-ray structures of all nine enzymes complexed with a wide variety of ligands. This is not achieved yet, but the next sections describe the progress being made towards this goal.

Glycosomal Enzymes

Essential steps to be taken in the first phase of the project were the purification of the target enzymes and the determination of the amino acid sequences of glycosomal proteins. Next physico-chemical properties had to be determined, and single crystals to be grown in order to unravel three-dimensional structures by X-ray crystallography.

An elegant procedure has been developed at the Institute of Cellular and Molecular Pathology (ICP) in Brussels which allows rapid purification of all nine enzymes involved (37). This was followed by investigations of physico-chemical properties, the results of which are summarized in table 1. Kinetic studies which are currently being carried out in Brussels by Dr. Fred Opperdoes and co-workers are revealing interesting differences between glycosomal enzymes and their mammalian counterparts.

Some of the major differences observed so far are the following:

- (i) Hexokinase and phosphofructokinase have completely different molecular weights compared with their mammalian homologues (33);
- (ii) All enzymes, except phosphoglucose isomerase, have a very high isoelectric point (33);

Table 1. The nine enzymes occurring in the glycosome with number of subunits per oligomer, molecular weight per subunit and isoelectric point, pI (29).

Enzyme	Abbrev.	Number of subunits	MW (subunit)	pI
Hexokinase	HK	6	50,000	10.0
Phosphoglucose isomerase	PGI	2	62,000	7.5
6-phosphofructo kinase	PFK	4	50,000	8.7
Fructose-bisphosphate aldolase	ALDO	4	41,000	9.1
Glycerol-3-phosphate dehydrogenase	GDH	2	37,000	10.0
Glycerol kinase	GK	2	53,000	9.0
Triosephosphate isomerase	TIM	2	27,000	9.8
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	4	39,000	9.3
3-phosphoglycerate kinase	PGK	1	47,000	9.4

- (iii) The nucleotide specificity of glycosomal PGK is different from other PGK's (37);
- (iv) At low ionic strength and pH above 7 the activity of triosephosphate isomerase (TIM) is considerably less than that of TIM's from other organisms (38);
- (v) The affinity of GAPDH for the cofactor NAD is 5–10 times less in the trypanosome enzyme than in the human isozyme (Callens, M., Kuntz, D. and Opperdoes, F. R., submitted for publication).

Simultaneously with physico-chemical and kinetic characterization of glycosomal enzymes their genes have been subject of intense investigations. As a result *T. brucei* gene sequences are now known of glycosomal TIM (39), cytosolic and glycosomal PGK (40), glycosomal GAPDH (41), cytosolic GAPDH (42), glycosomal aldolase (43, 44) and glycosomal PGI (45).

Import of Enzymes into Glycosomes

Glycosomal enzymes are synthesized in the cytosol on free polysomes. The half-life of the mature enzyme in the cytosol is approximately 3 minutes (46, 47). Subsequently they are taken up by the glycosome without post-translational modification (46). The mechanism of import into glycosomes is unknown as yet. From a comparative study of amino acid sequences and inspection of approximate models of three glycosomal enzymes a hypothesis was proposed by Wierenga *et al.* (48). The basic idea was that two clusters of positive charges about 40 Å apart on the surface of glycosomal enzymes might constitute an important component of the import signal for glycosomes.

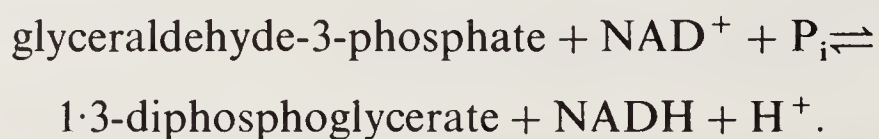
This hypothesis does not appear to hold up well in the light of new data which has become available. Recent gene sequence data from Swinkels and Borst (49) on glycosomal and cytosolic PGK's from the insect trypanosome *Crithidia fasciculata*, provided no evidence at all for positive clusters of charges being crucial differences between these two enzymes. Virtually the only difference in sequence of these PGK's

is the C-terminal extension of glycosomal PGK. This suggests that the import signal resides in the C-terminal extension of glycosomal PGK. Not all glycosomal enzymes have a C-terminal extension, however (41, 43–45, 48). Very recently, the gene sequence of glycosomal GAPDH from *Trypanosoma cruzi* (50) shows that one of the positive clusters suggested to be important for glycosomal import was no longer intact. Hence, the excess of positive charge which so prominently occurs in glycosomal enzymes from *Trypanosoma brucei* (48) may have other functions. Michels (51) and Opperdoes (52) suggested neutralization of negatively charged metabolites which are present in high concentrations in the glycosomes. Clearly, further research is required to establish the glycosomal import signal.

One final remark concerns potential possibilities to block import which would not depend on precise knowledge of import signals. In the case of import into mitochondria, studies of Schatz *et al.* (53) indicate that compounds with high affinity for the protein to be translocated prevent membrane passage. This is interpreted as evidence for at least partial unfolding during the import process. If this unfolding would also be essential for import into glycosomes, then every low molecular weight compound which binds very tightly in any of the clefts of a glycosomal enzyme might prevent uptake into the glycosome. If this were true then we would not be limited to compounds binding selectively to active sites or subunit interfaces, but could explore the entire surface of glycosomal proteins. It will obviously be of great interest to see if unfolding is indeed a feature of the glycosome import process. As this has not yet been established, we refrain from further discussions of glycosome targetting and turn to two glycosomal enzymes, which have been studied with respect to their three-dimensional structures.

Glycosomal Glyceraldehyde Phosphate Dehydrogenase

Glyceraldehyde phosphate dehydrogenase (GAPDH) is a tetramer with subunits of approximately 37,000 daltons. The enzyme catalyzes the following reaction:



The enzyme undergoes considerable conformational changes upon binding of NAD, as has been revealed by crystallographic studies on holo- and apo GAPDH from lobster and *Bacillus stearothermophilus* (54, 55).

The degree of amino acid sequence identity of lobster, *B. stearothermophilus* and human enzyme, with each other and with the glycosomal *T. brucei* enzyme (41) is shown in table 2. It is obvious that the *T. brucei* enzyme differs significantly from the two enzymes with known three-dimensional structure (54–56) so that an accurate structure determination of *T. brucei* GAPDH is required for our drug design purpose. A low resolution structure of human GAPDH (57) was available at the beginning of our studies with obvious limitations as to its accuracy. Hence we needed not only to determine the atomic structure of *T. brucei* GAPDH, but also to improve the accuracy of the human structure.

We are most grateful to Prof. H. Watson from Bristol, UK, for providing us with enzyme, crystals, data and initial coordinates which allowed us to obtain a quite

Table 2. Percentage of sequence identity of lobster *B. stearo-thermophilus*, human and *T. brucei* glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (41).

	Lobster	<i>B. stearo-thermophilus</i>	Human	<i>T. brucei</i> (glycosomal)
Lobster	100	53	70	57
<i>B. stearo-thermophilus</i>	53	100	52	54
Human	70	52	100	55
<i>T. brucei</i> (Glycosomal)	57	54	55	100

accurate structure of the human enzyme. Using one crystal, a 84% complete data set out to 2.4 Å resolution was collected on the FAST television detector diffractometer in Groningen (R. J. Read *et al.*, unpublished results). Refinement of the starting model, which has half a tetramer, or 80,000 daltons, in the asymmetric unit, has resulted in a structure which has a quite reasonable *R* factor of 22%, with good geometry and without water molecules. In particular near the substrate and the co-enzyme binding sites the currently available model provides a reliable reference point for assistance in the design of compounds which would block the functioning of glycosomal GAPDH but should leave the human enzyme largely unaffected.

The structure determination of glycosomal GAPDH is making good progress as well. Purified enzyme gave two crystal forms (58) both with quite large cell dimensions, rather small size and considerable radiation sensitivity. Only with the use of synchrotron radiation a 85% complete 2.8 Å resolution data set of form II crystals could be recorded on film. For the data collection a total of 16 crystals were used, and altogether 134 films were processed. After merging and scaling, a dataset of 57,050 unique reflections was obtained. Next, the molecular replacement method (59), starting from the *B. stearothermophilus* GAPDH structure, allowed a fairly straightforward initial solution of the *T. brucei* GAPDH form II structure, in spite of the considerable complexity of the subunit arrangement in these crystals. It appeared that the crystals contain one complete and one half tetramer in the asymmetric unit, which means that we are dealing with 6 subunits of 37,000 daltons each, or a total of ~ 220,000 daltons. Refinement of this crystal structure is still underway. The current model has an *R* factor of 44% and obviously still needs careful rebuilding at several positions.

Due to scarcity of protein material too few crystals were left to do either soaking studies or to determine the structure of crystal form I using the ‘classical’ rotation method. Since the morphology of the crystals did not allow distinction of forms I or II we decided to see if the application of the ‘Laue method’ would yield structures of sufficient accuracy for our design purposes, while keeping the need for crystals to a minimum. It turned out that this led to a quite unconventional structure determination of crystal form I of glycosomal GAPDH from *T. brucei*.

Only two crystals were required to collect four Laue film packs with a total exposure time of 12.5 seconds. The film packs of six films each were processed using the CCP4

Laue film processing software. This eventually led to a dataset which was 37% complete between 7 and 3.2 Å resolution, containing 22,022 unique reflections.

Cell dimensions of crystal form I are $a = 135.5$ Å, $b = 256.3$ Å, $c = 114.9$ Å, space group $P2_12_12_1$, which allows for a large number of subunits in the cell. Possibilities range from $1\frac{1}{2}$ to 3 tetramers, or 220,000 to 440,000 daltons per asymmetric unit. Starting from the well-refined *B. stearotherophilus* GAPDH structure (54, 56; see figure 5) it was possible to solve this molecular replacement problem, resulting in an

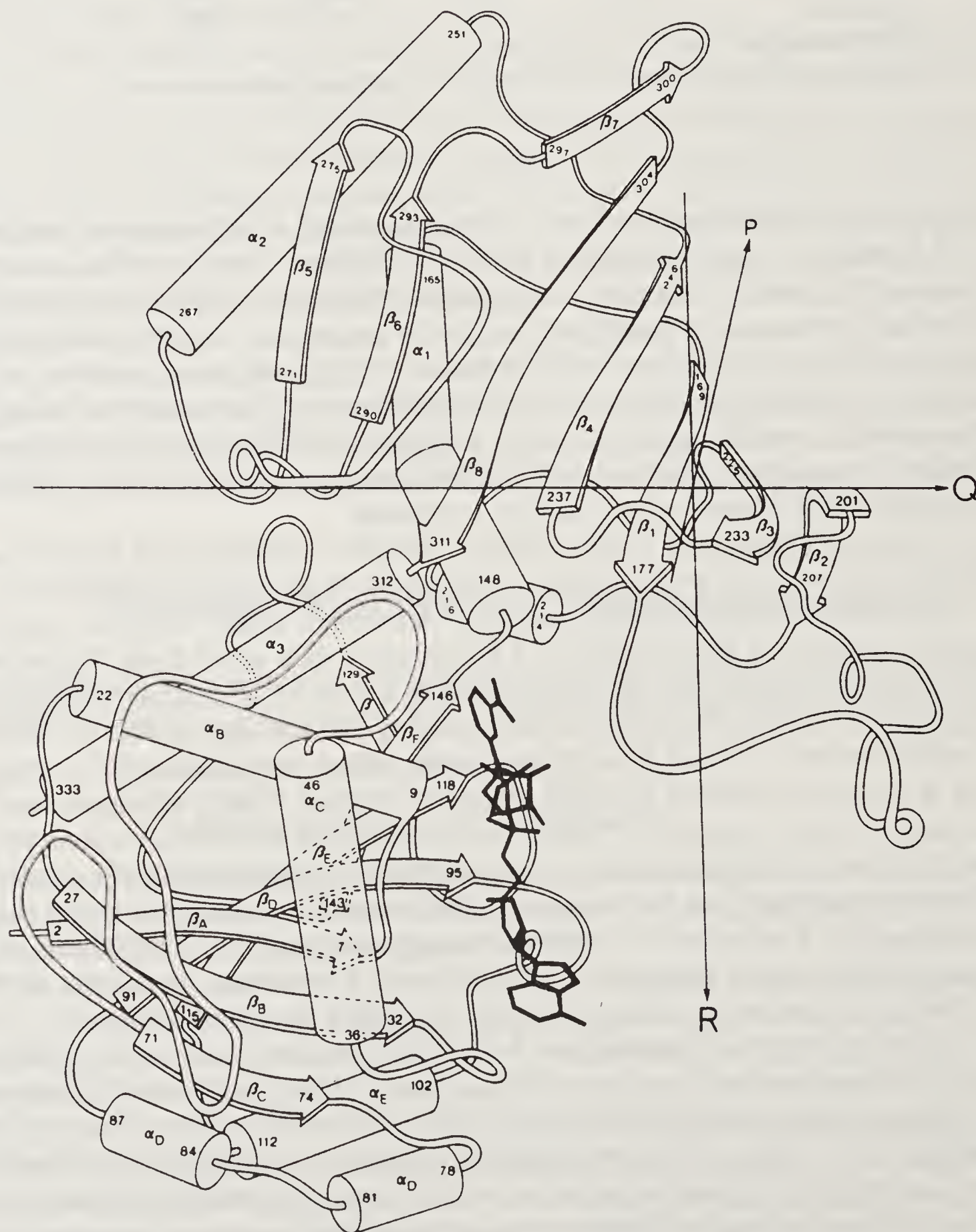


Figure 5. Schematic drawing of *B. stearotherophilus* glyceraldehyde phosphate dehydrogenase. Cylinders indicate α -helices, arrows β -strands. The NAD is shown in heavy lines. The twofold axes indicated with P, Q and R relate the subunit shown to the other members of the tetrameric molecule (Reproduced, with permission, from Biesecker *et al.* (80))

initial model of $1\frac{1}{2}$ tetramers, or 6 subunits, per asymmetric unit. The six-fold redundancy was subsequently used for:

- (i) Density averaging;
- (ii) Introduction of calculated phases and amplitudes for unobserved reflections; and,
- (iii) Constrained refinement imposing six-fold symmetry on the atomic coordinates using simulated annealing procedures. The resultant model needs still further rebuilding and refinement, but a reasonably accurate model of also this crystal form of gGAPDH has now been obtained (F. M. D. Vellieux, J. Hajdu, C. L. M. J. Verlinde, R. J. Read, T. Greenough, J. Littlechild, H. Watson and W. G. J. Hol, in preparation). So, for future experiments structures of both crystal forms of the glycosomal enzyme, as well as of the human enzyme, are now available.

Exploitation of this structural information is only just starting. Fortunately, the sequence of glycosomal GAPDH has been determined within the framework of this collaborative project by Michels *et al.* (41). Inspection of sequences and structural models revealed most interesting differences near the adenine binding pocket of the co-factor NAD (table 3; figure 6). For instance, in the “back” of this pocket the asparagines 33 and 200 appear to be hydrogen bonded to each other, both in the human as well as in the *B. stearothermophilus* enzymes. In glycosomal GAPDH, however, Asn 33 is replaced by a valine. Hence, the characteristics of the binding pocket change significantly. Initial and preliminary suggestions are that modified adenines, adenosines, AMP or ADP or molecules with appropriate ‘variants’ of the adenine ring might be better inhibitors of *T. brucei* glycosomal GAPDH than of the human enzyme. Other modifications, involving for instance the 2' and 3' hydroxyls of the adenine ribose, might also lead to selective inhibition of the trypanosomal enzyme.

While these structure determinations and modelling efforts were proceeding, the good news reached us that *T. brucei* glycosomal GAPDH was successfully over-expressed in *E. coli* in Brussels by Dr. P. A. Michels and co-workers. This is a dramatic change: the target enzyme was so far only available in small amounts, whereas now a virtual “unlimited” supply of protein material is available. Hopefully, new crystals can be obtained using the recombinant material, which should lead to the design of

Table 3. Residues near the adenine moiety of NAD differing between human and glycosomal GAPDH of *Trypanosoma brucei*.

<i>T. brucei</i>	Human
Asn 7	Asp 7
Val 36	Asn 33
Asn 39	Phe 36
Ala 89	Glu 78
Asn 92	Asp 80
Leu 112	Val 100

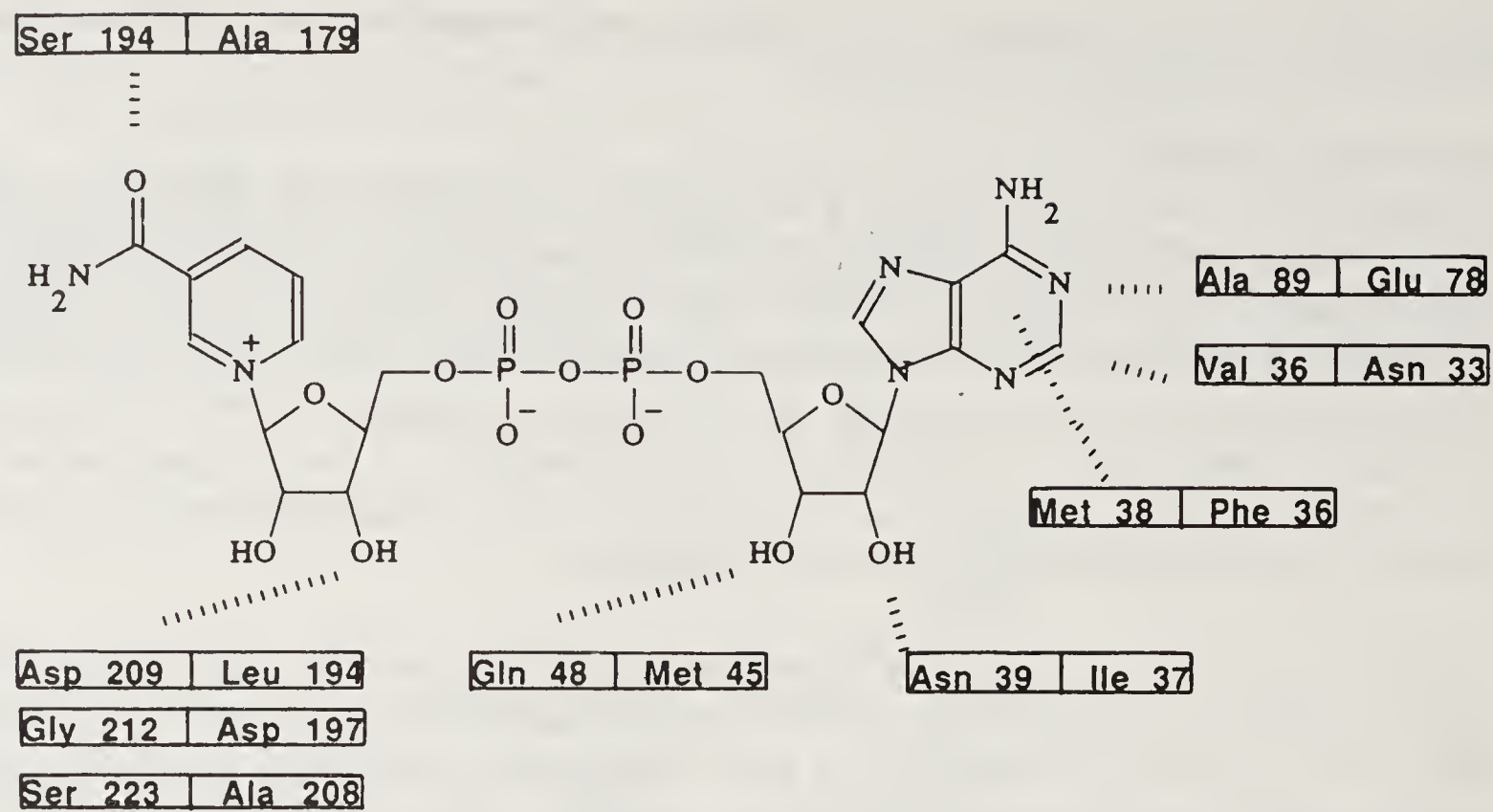


Figure 6. Indication of amino acids near NAD bound to GAPDH which are different between the *T. brucei* glycosomal enzyme and the human enzyme. The left residue in each box is the residue in GAPDH, the right one the equivalent residue in human GAPDH.

selective inhibitors of glycosomal *T. brucei* GAPDH by the cyclic procedure depicted in figure 1.

Glycosomal Triosephosphate Isomerase

General Features and Amino Acid Sequence

Triosephosphate isomerase (TIM) is a well characterized glycolytic enzyme. Its function in glycolysis is indicated in figure 4. The enzyme catalyzes the reaction shown in figure 7. TIM consists of a dimer with subunits of ~ 250 amino acid residues each, resulting in a total molecular weight of ~ 55,000. The enzyme is only fully active as a dimer

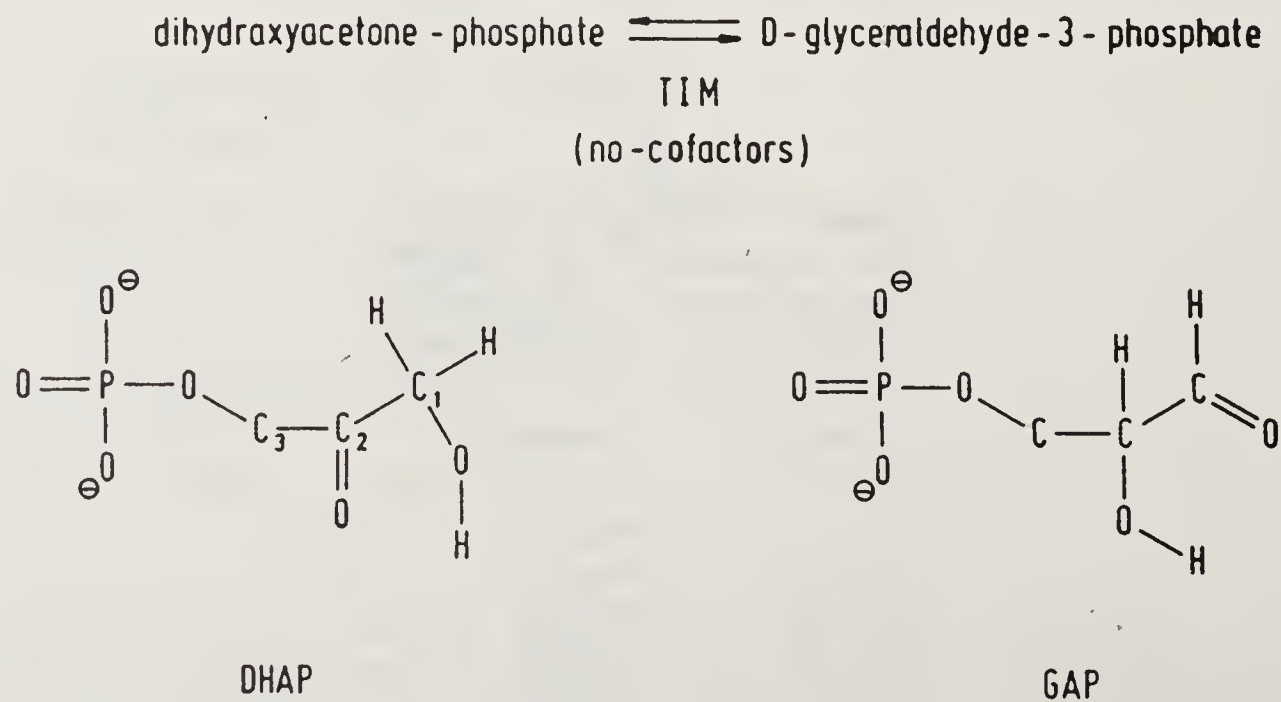


Figure 7. Overall reaction and structures of the substrates of triosephosphate isomerase (TIM).

(60), but there is no evidence for cooperativity. During catalysis no co-factors are required. A wealth of kinetic evidence has led to the conclusion that triosephosphate isomerase is a very efficient enzyme indeed (61). Human beings missing the enzyme suffer from very poor health and die at an early age (62). The trypanosome, which is more dependent on glycolysis than humans, will depend even more on a proper functioning TIM. For this reason inhibitors of *T. brucei* TIM which leave the activity of human TIM largely unaffected are of great interest for the treatment of sleeping sickness.

The three-dimensional structure of chicken triosephosphate isomerase has been determined in the mid-seventies by the Oxford group (63, 64). The yeast enzyme has also been studied in detail (65, 66). The structure exhibits pseudo eightfold symmetry with 8 β -strands forming a central parallel β -barrel surrounded by 8 α -helices, linked together by a number of loops. A schematic view of a single subunit is depicted in figure 8.

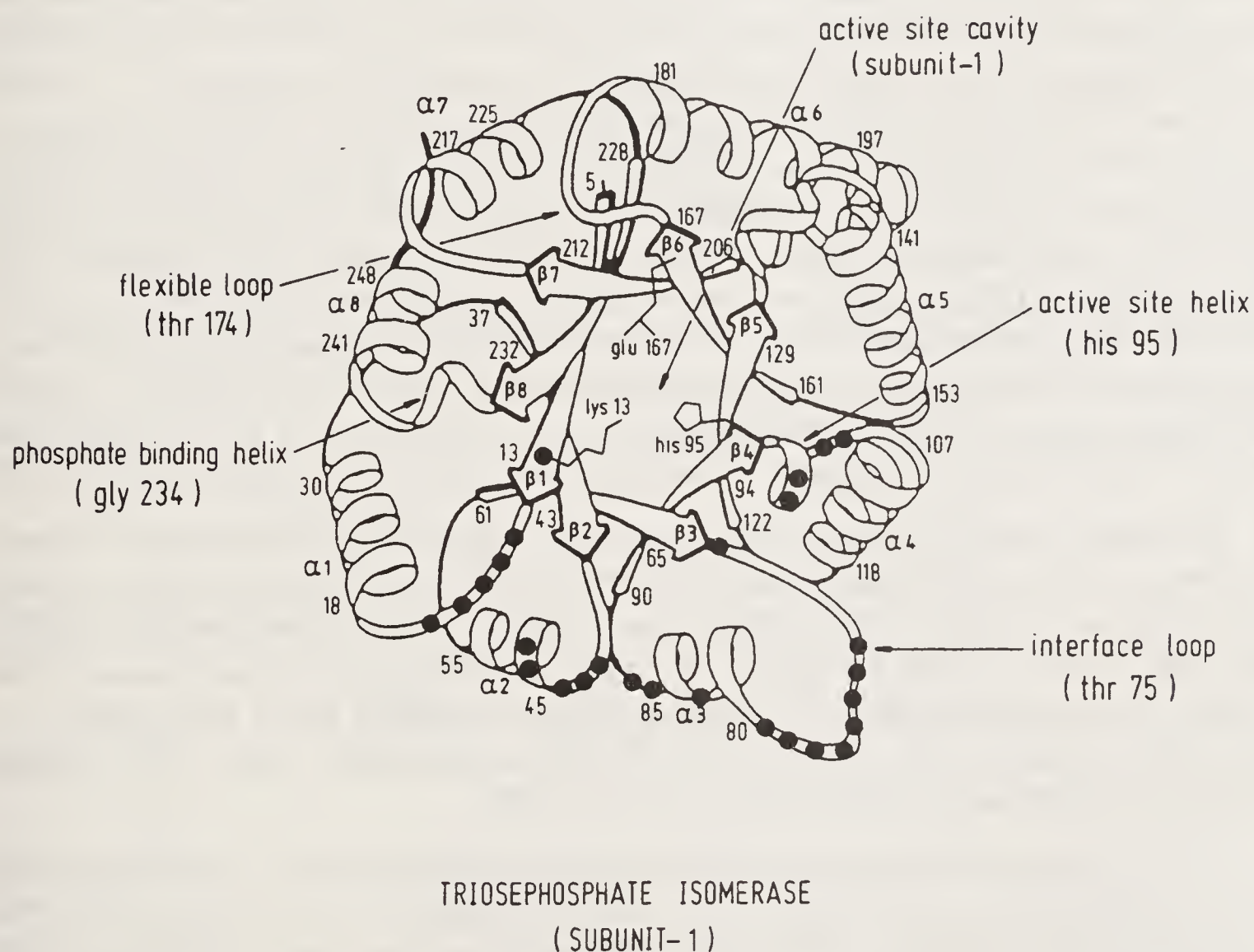


Figure 8. A schematic drawing of subunit-1 of the *T. brucei* TIM dimer. The approximate positions of the active site residues Lys 13, His 95 and Glu 167 are shown. The black dots, emphasizing the interface region, indicate the 27 residues with atoms which are closer than 3.5 Å from atoms of the other subunit (68). The “interface loop” (residues 72–79) points towards the active site of subunit-2. The “flexible loop” (167–180) closes over the active site after substrate binding (64–66, 71, 72). The “active site helix” is formed by residues 95–103. The “phosphate binding helix” is formed by residues 234–239. AYY shows the location of residues Ala 100–Tyr 101–Tyr 102, these residues have the sequence His–Val–Phe in human TIM and constitute amino acid residue differences between human and parasite enzyme occurring relatively close to the active site. Residues of subunit-1 are numbered 2–250, those of subunit-2 302–550.

General features of triosephosphate isomerase (63–72) are indicated in figure 8 of the *T. brucei* enzyme. They include:

- three residues in the active site cavity which are directly interacting with the substrate: Lys 13, His 95 and Glu 167;
- the position of His 95 situated at the N-terminus of a short “active site helix” comprising residues 95–105;
- the phosphate moiety of the substrate interacting with peptide units of the N-terminus of a “phosphate binding helix” (67);
- at the end of β_6 residues 167–180 form a so-called “flexible loop” which covers the active site during catalysis (64, 69, 70);
- helices α_2 and α_3 plus the loop between β_1 and α_1 plus the loop between β_2 and α_2 are involved in contacts with the other subunit of the dimer, while residues 74–77, between β_3 and α_3 , form the so-called “interface loop” which protrudes into the active site of the other subunit (68).

When the structure of yeast TIM became available (65–66) it became clear that the conformations of the chicken and yeast enzymes were essentially the same, despite a 53% difference in the amino acid sequences. Recently, the sequence of *T. brucei* TIM has been derived from the available gene sequence (40). Also the sequence of human TIM is known (62, 73). An alignment of all four sequences is given in figure 9. The *T. brucei* sequence is approximately 50% identical with other known sequences, both eukaryotic and prokaryotic. The human TIM sequence is 88% identical with that of chicken TIM so that only minor alterations are required to obtain a quite accurate three-dimensional model of human TIM. For trypanosomal TIM this is clearly not the case and a structure determination has been carried out. Before describing aspects of the crystallographic work, comparison of the *T. brucei* and human TIM sequences in the light of the chicken and yeast structures showed that (68):

- (i) The active sites differ very little in sequence between the human and *T. brucei* enzyme: $\sim 85\%$ of all residues within 10 Å from His 95 are identical. Surface residues near to the active centre which differ between parasite and host enzyme are residues 100–102. These are Ala–Tyr–Tyr in *T. brucei* and His–Val–Phe in human. These residues are approximately 10 Å removed from the essential active site atoms;
- (ii) Quite in contrast, the interface residues differ substantially. Only $\sim 55\%$ of these residues are identical in human and *T. brucei* TIM.

In view of these limited sequence differences near the active site it is understandable that few differences exist between the kinetic properties of *T. brucei* and other triosephosphate isomerases. The kinetic differences at low ionic strength and high pH (38) may be due to differences in the subunit interface (68).

Crystal Structure of Glycosomal Triosephosphate Isomerase

In spite of the fact that TIM is the least abundant glycolytic enzyme in the glycosome (29), it was the first to give good quality crystals and hence the first glycosomal enzyme for which a structure was determined. In fact it is the first enzyme from any tropical parasite with known three-dimensional structure (68).

Crystals were obtained from solutions with a high ammonium sulphate concentration and crystallized with a complete dimer in the asymmetric unit. Using five crystals



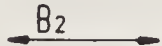

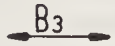
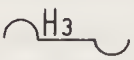
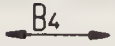
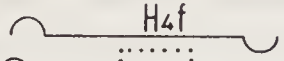

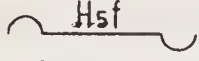
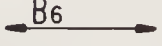


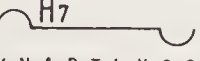
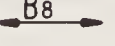
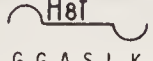

					
T. brucei TIM	(2) S K P	Q P I A A A N W ^(K)	C N G S	Q Q S L S E L I D L F N S	(30)
Human TIM	(1) A P S R	K F F V G G . . .	M . . R	K . . . G . . . G T L . A	(30)
Chicken TIM	(2) A P R	K F F V G G . . .	M . . D	K K . . G . . . H T L . G	(30)
Yeast TIM	(2) A R	T F F V G G . F .	L . . .	K . . I K . I V E R L . T	(29)
					
T. brucei TIM	(31) T S I N H D	V Q C V V A S	T	F V H L A M T K E R L	(55)
Human TIM	(31) A K V P A .	T E V . C A P	P	T A Y I D F A R Q K .	(55)
Chicken TIM	(31) A K L S A .	T E V . C G A	P	S I Y . D F A R Q K .	(55)
Yeast TIM	(30) A . . P E N	. E V . I C P	P	A T Y . I D Y S V S L V	(54)
					
T. brucei TIM	(56) S H P K F	V I A A Q	N A I A K S - G A F T G E V S	L P I L K D	(85)
Human TIM	(56) D P - . I	A V C Y K V T N I .	P G M I . .	(85)
Chicken TIM	(56) D A - . I	G V C Y K V P K I .	P A M I . .	(85)
Yeast TIM	(55) K K . Q V	T V G Y L . A S N .	V D Q I . .	(85)
					
T. brucei TIM	(86) F G V N	W I V L G	^(H) S E R R A Y Y : G E T N	E I V A D K V A A A V A	(118)
Human TIM	(86) C . A T	. V H V F . . S D	. L I G Q . . . H . L .	(118)
Chicken TIM	(86) I . A A	. V I H V F . . S D	. L I G Q . . . H . L .	(118)
Yeast TIM	(86) V . A K	. V I S . F H . D D	K F I . . . T K F . L G	(118)
					
T. brucei TIM	(119) S G F	M V I A C I G E	T L Q E R E S G R T A	V V V L T Q I A A I A K K	(153)
Human TIM	(119) E . L	G	K . D . . . A . I . E	K . . F E . T K V . . D N	(153)
Chicken TIM	(119) E . L	G	K . D . . . A . I . E	K . . F E . T K . . . D N	(153)
Yeast TIM	(119) Q . V	G . . L E . K K A . K . L	D . . E R . L N . V L E E	(153)
					
T. brucei TIM	(154) L K K A D W A	K V V I A Y ^(E)	P V W A I G T G K V A T P	Q Q A Q E A H A L I R S W V S S K	(197)
Human TIM	(154) V . - - . S	. . . L T V . E K L . G . L K . N	(195)
Chicken TIM	(154) V . - - . S	. . . L T V . E K L . G . L K . H	(195)
Yeast TIM	(154) V . - - . T	N . . V L A . . .	E D . . D I . . S . . K F L A . .	(195)
					
T. brucei TIM	(198) I G A D V R G E	L R I L Y G G	S V N G	K N A R T L Y Q Q	(225)
Human TIM	(196) V S D A . A Q S	T . . I T .	A T C K E . A S .	(223)
Chicken TIM	(196) V S D A . A Q S	T . . I T .	G . C K E . A S .	(223)
Yeast TIM	(196) L . D K A A S A . .	S . . V . F K D K	(223)
					
T. brucei TIM	(226) R D	V N G F L	V G G A S L K P	E F V D I I K A	(248)
Human TIM	(224) P .	. D N .	(246)
Chicken TIM	(224) H .	. D N .	(246)
Yeast TIM	(224) A .	. D N S	(246)
					
T. brucei TIM	(249) T Q				
Human TIM	(247) K .				
Chicken TIM	(247) K H				
Yeast TIM	(247) R N				

Figure 9. Amino acid sequence alignment of human (62,73) and *T. brucei* (40) triosephosphate isomerase, along with the two enzymes with known three-dimensional structure, those of chicken and yeast (81,82). Dots indicate that a residue is identical to the *T. brucei* residue. On each line a β -strand, a loop and an α -helix are shown, indicated by B1–B8 for the strands and H1–H8 for the helices. H4f, H5f and H8f are helices occurring in loop regions at the C-terminal ends of the β -strands, near the active site. H7b is a small helix at the other side of the subunit (see also figure 8). The important amino acid differences near the active site between human and *T. brucei* TIM (see text), AYY, residues 100–102, are in a box of dotted lines. The catalytically important residues Lys 13, His 95 and Glu 167 are encircled.

grown from the first batch of 0.6 mg purified protein a complete dataset of $\sim 12,000$ unique reflections to a resolution of 2.4 \AA could be obtained. Starting from the chicken TIM structure, rotation and translation functions (59) gave the orientation and position of the model structure which was subsequently refined by restrained Least Squares and Computer Graphics techniques (68). In order to obtain a very accurate structure for our molecular modelling studies, an additional 1.83 \AA data set was collected on films using the intense X-ray beam of the Daresbury synchrotron. The 115 film packs collected from 11 crystals were processed, giving a total of $\sim 230,000$ measured intensities. This gave a set of $\sim 39,000$ unique reflections. The refinement of the high resolution structure of gTIM has now been completed with an R factor of 18.3%, good geometry of the model, inclusion of 279 water molecules, two sulphate ions and one dithiothreitol molecule (R. K. Wierenga *et al.*, to be published).

This accurate structure has proved to be an exciting starting point to improve our knowledge of the properties of *T. brucei* TIM by crystallographic binding as well as by molecular modelling studies. These investigations are currently in full swing and some examples of the results obtained will be given in the next few sections.

Crystallographic Binding Studies

Protein crystals consist $\sim 50\%$ of mother liquor. This allows small molecules to diffuse through solvent channels and reach their binding site in the crystal. This approach can fail for a number of reasons such as: (i) the binding site is blocked by a neighbouring protein molecule; (ii) ions or molecules in the crystallization medium occupy the binding site; (iii) conformational changes, occurring upon binding of the small molecule, destroy the crystal lattice. In the case of *T. brucei* TIM all three of these factors appear to play a role as we now know:

- (i) The active site of "subunit-1" in the asymmetric unit is in contact with a neighbouring molecule and hence largely inaccessible for medium-affinity inhibitors;
- (ii) The entrance of the active site of "subunit-2" is occupied by a sulphate ion. Sulphate has a K_i of $\sim 5 \text{ mM}$ (38). Since the crystallization medium contains 2.4 M ammonium sulphate, straightforward inhibitor binding studies by soaking experiments are made infeasible;
- (iii) Major conformational changes occur near the active site upon substrate and inhibitor binding as reported for chicken and yeast TIM (63–66, 69) and are observed repeatedly in our own studies on *T. brucei* TIM (70–72).

This accumulation of problems was solved – at least for medium affinity inhibitors – by transferring *T. brucei* TIM crystals to another mother liquor. Hanging drop equilibration experiments were used to establish potentially suitable concentrations of PEG solutions (74) which could substitute for ammonium sulphate. Eventually a specific protocol was obtained which allowed transfer of the *T. brucei* TIM crystals from 2.4 M ammonium sulphate to 44% (w/w) PEG 6000 solutions without loss of their diffraction power (74). This breakthrough has allowed us to study crystallographically the mode of binding of inhibitors with inhibition constants down to $500 \mu\text{M}$. It turned out that use of an inhibitor with a K_i of $27 \mu\text{M}$, i.e. 2-phosphoglycollate, led to a deterioration of crystal quality, probably due to conformational changes near the active site of "subunit-1".

This is clearly still not an ideal situation because we are interested in the study of inhibitors with K_i 's much lower than $27\ \mu\text{M}$. Fortunately, very recently, the gene of *T. brucei* TIM has been overexpressed in Brussels by Dr. K. Pratt and Dr. P. A. M. Michels. This allows large-scale co-crystallization experiments which was not really feasible with the small amounts of enzyme which so far had to be obtained from trypanosomes grown in rats. Such co-crystallizations have now been initiated.

The study of the binding mode of a variety of inhibitors by means of soaking into the 'transferred' crystals has, however, already given us precious insight as to the surprising adaptability of TIM upon binding of a number of related compounds.

'Sulphate free' T. brucei TIM: Sometimes experiments fail to give an answer to the question addressed, but nevertheless provide useful information. This was the case in our attempts to unravel the binding mode of 1-amino naphthalene-4,6,8-trisulphonic acid (ANTS). This compound was kindly provided by Dr. Nickel in Bonn and the inhibition constant of *T. brucei* TIM was found to be $0.3\ \text{mM}$ (A. M. Lambeir & F. R. Opperdoes, personal communication). *T. brucei* TIM crystals were transferred to a 44% (w/w) PEG 6000 solution, as described above, containing also $2\ \text{mM}$ ANTS. Data out to $2.8\ \text{\AA}$ resolution were collected on the FAST area detector and the structure was determined starting from the well-refined model of TIM in $2.4\ \text{M}$ ammonium sulphate. Upon application of molecular dynamics refinement techniques, the "ANTS model" eventually reached an R factor of 13.9% with good geometry and 30 water molecules (70). No inhibitor was visible in the electron density map, however. This is quite puzzling, in particular since we have elucidated the binding mode of several other inhibitors successfully. We refrain here from trying to explain why the inhibitor did not bind but, instead, report briefly the results which were obtained by this study.

It appeared that "subunit-2" did not contain a sulphate ion any more. This resulted in a number of conformational changes the most important of which was a major repositioning of the "flexible loop" near the active site. This loop, comprising residues 167–180, is closed in the presence of sulphate and phosphate containing compounds (71, 72), but turned out to adopt an open conformation in the absence of sulphate. This conformational change causes the C^α atom of residue 475 (i.e. residue 175 of the 'second' subunit) at the top of the loop to move $6.9\ \text{\AA}$ towards the open conformation. As can be seen in table 4, the conformation of this loop in the presence of phosphate-containing substrate analogues is slightly different from the conformation in the presence of sulphate. The main point, however, is a major conformational change which occurs during the catalytic process.

The loop closure may have several functions. One of these is avoiding dephosphorylation as a side reaction during the catalytic process (75). Another one might be to bring active site residues into the optimal position with respect to the substrate. This may especially be true for Glu 167. This residue can adopt several conformations and is located at the very beginning of the flexible loop. Leaving mechanistic implications aside, this study provided us with a picture of the active site of "subunit-2" in the 'open' conformation (69, 70) as occurs in the absence of substrates and inhibitors. Although it has so far not yet been done, this 'open' conformation provides an alternative starting point for modelling studies to the 'closed' conformation which has been investigated in detail as we see in the next sections.

Table 4. Conformational changes of the flexible loop observed in *T. brucei* TIM (70, 71, Wierenga *et al.*, in preparation).

Residue number in 'subunit 2'	Difference of C α position between the 'open' and the 'closed' conformation (Å)			
	3PP	G3P	3PGA	Sulphate
Pro 468	1.0	0.8	1.1	0.8
Val 469	0.9	0.7	0.9	0.7
Trp 470	2.6	2.7	2.7	2.3
Ala 471	2.9	2.9	3.0	2.8
Ile 472	3.2	3.1	3.0	2.7
Gly 473	5.6	5.7	5.4	4.9
Thr 474	6.1	6.3	6.1	5.7
Gly 475	6.8	7.1	7.1	6.9
Lys 476	4.6	4.6	4.6	4.8
Val 477	4.8	4.7	4.6	4.9
Ala 478	1.7	1.9	1.7	1.5

The difference between sulphate and monohydrogen phosphate binding by T. brucei TIM: The active site pocket of TIM is deeply buried in the central part of the molecule (figure 8). As mentioned earlier, amino acid differences between trypanosomal and human TIM useful for the design of selective inhibitors are located quite far from the active site, e.g. the tripeptide 100–102, which is AYY in *T. brucei* TIM and HVF in human TIM, is located more than 12 Å away from the phosphate binding site at the entrance of the active site pocket (figure 8; table 5). Hence we are interested in elucidating the binding mode of phosphate containing molecules which bind with their 'non-phosphate' moiety *outside* the active site pocket. Modelling studies suggested that this might be the case for ribose-5-phosphate (R5P). M. Callens and F. Opperdoes (personal communication) determined that the K_i of this compound was ~ 5 mM.

For our crystallographic studies, crystals were transferred to 44% (w/w) PEG 6000 (74) containing 15 mM ribose-5-phosphate. Data were collected on the FAST area detector yielding 14,592 unique reflections which is 88% of the data out to 2.4 Å resolution. Refinement resulted in a molecular model with an R factor of 15.0%, with

Table 5. Distances (Å) between the phosphor atom of glycerol-3-phosphate complexed to *T. brucei* TIM and specific atoms of the tripeptide 100–102 Ala Tyr Tyr (numbered 400–402 in the 'second' subunit).

C α (Ala 400)	15.6
C β (Ala 400)	14.4
C α (Tyr 401)	16.7
OH (Tyr 401)	12.1
C α (Tyr 402)	19.0
OH (Tyr 402)	16.3

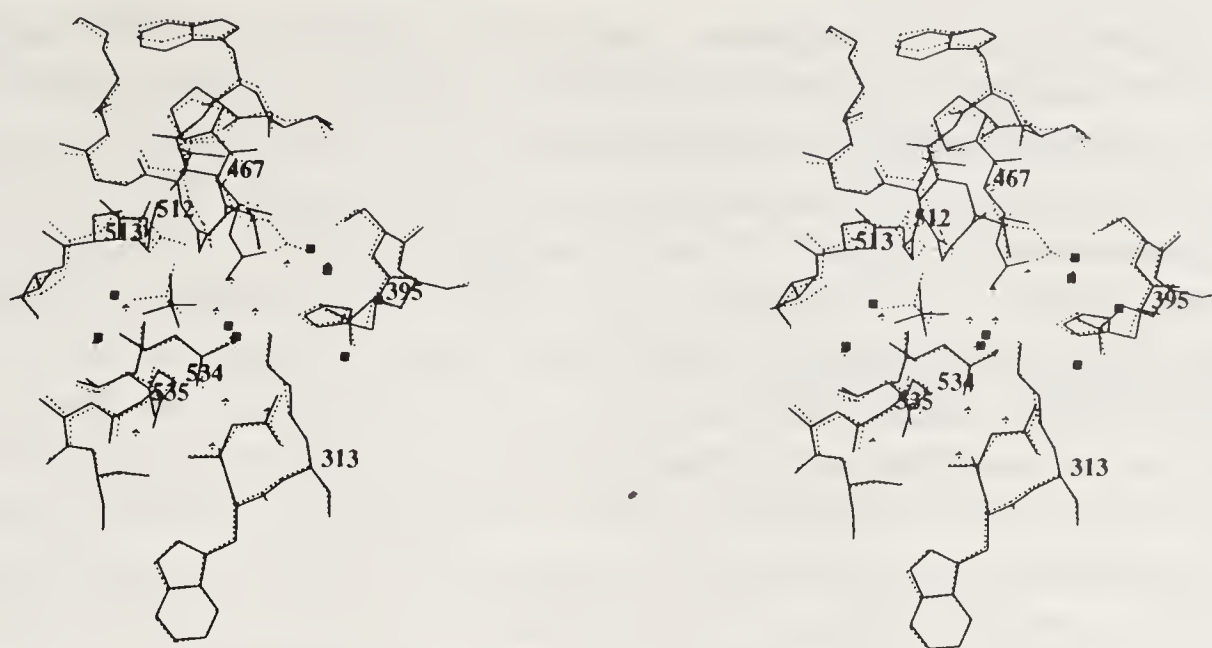


Figure 10. Stereo drawing of the active site of glycosomal TIM from *T. brucei* complexed with sulphate (dashed lines and waters indicated with crosses) and complexed with phosphate (as deduced from the ribose-5-phosphate soaking experiment in solid lines and waters indicated with squares).

good geometry, 156 water molecules, and one monohydrogen phosphate ion bound to the active site of “subunit-2”.

No ribose moiety was visible in the electron density map, almost certainly due to hydrolysis of the inhibitor molecule. This would explain why a phosphate ion is observed at the entrance of the active site pocket of “subunit 2”. This result gave us the opportunity to compare the binding mode of the divalent sulphate ion with the divalent monohydrogen phosphate ion (Verlinde *et al.*, submitted). This turned out to be most intriguing.

The main result is that the sulphate position in the 1.83 Å ammonium sulphate structure is about 1 Å further ‘out’ of the active site pocket than the phosphate position. At the same time, the flexible loop is ~ 0.5 Å less well closed in the ‘sulphate complex’ than in the ‘monohydrogen phosphate complex’. Also, the water molecules within the active site pocket seem to be arranged in a distinctly different manner. And, finally, the catalytically essential Glu 467 adopts a different side chain conformation (figure 10).

A survey of the Cambridge structural data base reveals that the differences in size between SO_4^{2-} and HPO_4^{2-} are less than 0.1 Å. Charges on the oxygen atoms are quite similar for the two ions. Hence, the presence of the extra hydrogen atom on one of the oxygens of HPO_4^{2-} is likely to be the cause of the differences observed in the active site. It remains probably quite a challenge for molecular modellers to arrive by theory alone at this difference of 1 Å in positions of ions which are so similar, and also have virtually identical inhibition constants.

One important message of this study is that subtle features have distinct effects on the binding mode of related molecules to the same enzyme. Another message seems to be that divalent ions can possibly be bound in a variety of positions ranging by 1 Å between the “phosphor” and the “sulphur” sites. However, as we shall see in the next sections, studies on organic phosphates and phosphonates have so far not supported this hypothetical variability in phosphor position.

Differences between a phosphate and a phosphonate: Starting from the proposed transition state (76), several transition state analogues of TIM have been investigated

(77, 78). Wolfenden and co-workers studied the inhibition of TIM activity by a transition state analogue, 2-phosphoglycolate (77) (figure 11). This compound turned out to be an excellent inhibitor indeed, having a K_i of 0.027 mM for trypanosomal TIM (38). The structure of 2-phosphoglycolate led the late Dr. Alan Horn to suggest 3-phosphonopropionate (figure 10) as a potential inhibitor. This compound was bought from Aldrich and appeared to be a much less powerful inhibitor than 2-phosphoglycolate, the K_i being 27 mM (A.-M. Lambeir, unpublished observations).

In order to obtain insight into the reasons for this difference in affinity between a phosphate and a phosphonate of a factor of 10^3 – the difference being the change of an oxygen atom into a methylene group – crystallographic binding studies were initiated. Crystal packing effects made it impossible to study 2-phosphoglycolate, as explained in a previous section. The structure of *T. brucei* TIM complexed with 3-phosphono propionate (3PP) could, however, be determined.

The glycosomal TIM crystals were transferred, by the usual procedure (74), to a PEG solution containing 100 mM 3PP. Crystals were of limited quality so only a data set to 2.8 Å resolution could be obtained. The 10,879 unique reflections represent 75.7% of the data out to this resolution limit. The crystallographic R factor of the 3PP–TIM complex was 12.5% after refinement for a model containing 9 ligand atoms, 38 water molecules and 3778 non-hydrogen protein atoms (71).

The 3PP molecule is bound to the active site pocket with each oxygen atom involved in a hydrogen bond, one of which is a hydrogen bond to the side chain of Glu 467. This residue is probably protonated, in analogy with the presumed protonation state of Glu 467 after binding of 2-phosphoglycolate to yeast TIM (66). The main points we like to focus at here is that the phosphate of 3PP is within 0.2 Å of the phosphate of monohydrogen phosphate discussed in the previous section. Also, the flexible loop is ‘fully closed’ as in the complex with monohydrogen phosphate – not ‘almost closed’ as in the complex with sulphate. Clearly, the low affinity of 3PP for *T. brucei* TIM is *neither* due to lack of closure of the flexible loop *nor* due to an unusual position of the phosphate moiety. It is also not due to the fact that the ether phosphate oxygen of 2-phosphoglycolate is interacting with a protein hydrogen bond donor – it is *not*, according to the study of Lolis and Petsko (66) TIM as well as to our own studies on inhibitors like glycerol-3-phosphate, bound to *T. brucei* TIM (71). Replacing this ether phosphate oxygen by a CH_2 moiety does not therefore result in an unsatisfied protein hydrogen bond donor.

Hence, the reasons for the factor of 10^3 difference in K_i of 2-phosphoglycolate and 3-phosphonopropionate might stem from:

- (i) Less favourable torsion angles when 3PP is bound to TIM, compared with 2PGH;
 - (ii) Differences in interactions with water molecules of the two inhibitors;
 - (iii) A difference in pK of phosphate and phosphonate groups (see e.g. also ref. 79).
- In view of the absence of a structure of 2-phosphoglycolate with *T. brucei* TIM and the somewhat limited resolution of the complex with 3PP it is probably best not to carry on here with a long discussion of possible explanations. Rather we turn to other experimental results.

Binding modes of glycerol-3-phosphate and 3-phosphoglycerate: In order to increase further our insight as to the properties of *T. brucei* TIM when binding low molecular weight compounds, we studied the inhibitors glycerol-3-phosphate (G3P) and

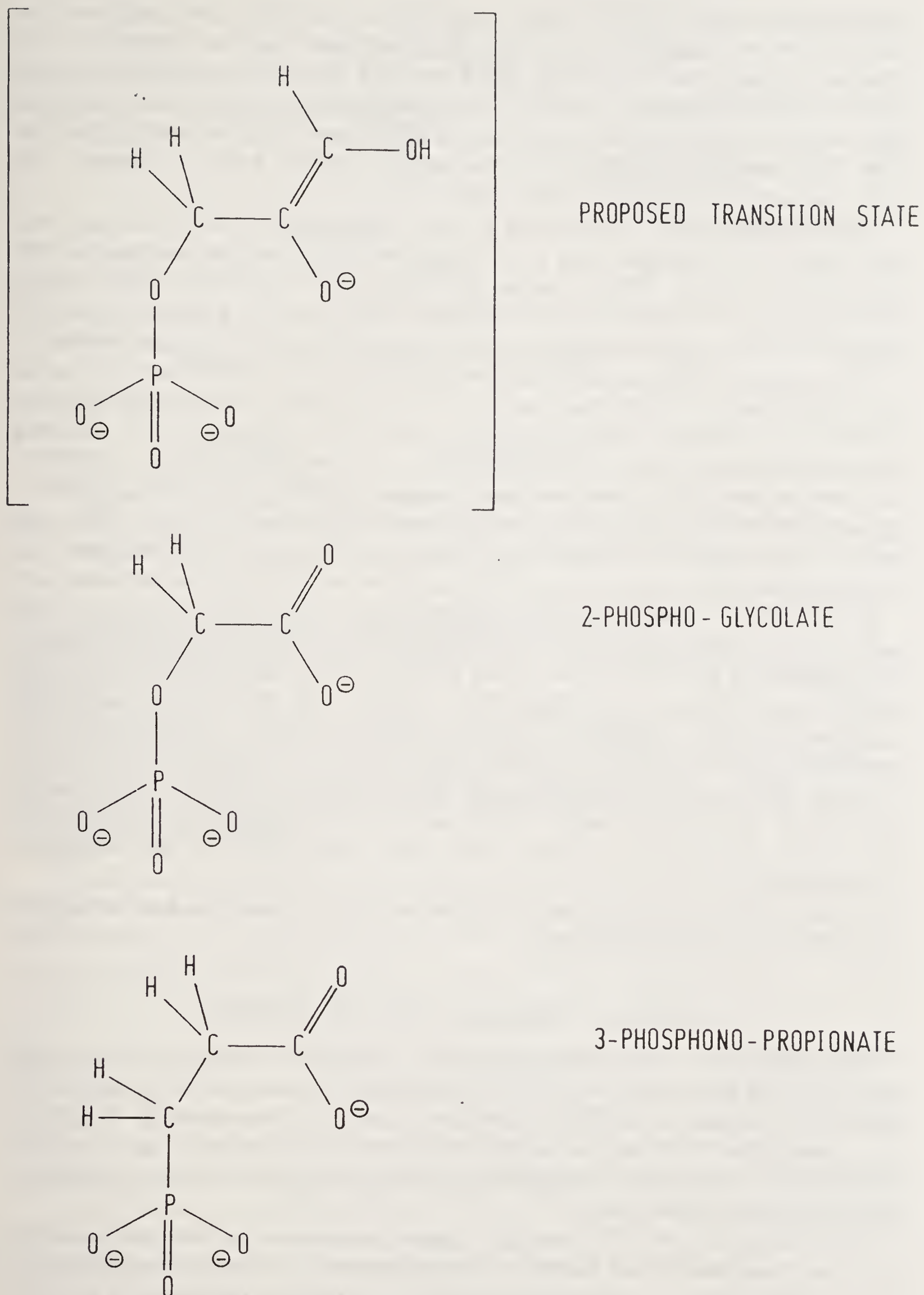


Figure 11. Molecular formulae of (i) the proposed transition state, (ii) a transition state analogue 2-phosphoglycolate ($K_i = 0.027 \text{ mM}$) and (iii) an inhibitor, 3-phosphono-propionate ($K_i = 27 \text{ mM}$), of triosephosphate isomerase. Only one atom is different in the latter two compounds, due to the replacement of an oxygen atom in 2-phosphoglycolate by a methyl group in 3-phosphono propionic acid.

3-phosphoglycerate (3PGA) in complex with the enzyme. Both these inhibitors have one atom more than 3PP when going along the chain from the phosphor to the terminal oxygen atom. In addition, 3PGA and G3P have one extra hydroxyl group at the 2-position compared with 3PP. 3-Phosphoglycerate and glycerol-3-phosphate differ merely by changing the carboxylate of 3PGA into a hydroxy-methyl group in G3P. This removes a charge near to the catalytic Glu 467 from the complex with G3P compared to the complex with 3PGA.

The three-dimensional structures of *T. brucei* TIM with G3P and 3PGA have been determined in an analogous way to that described in the previous section for 3PP. The *R* factors for the 3PGA complex after refinement is 14.0% for 16,813 unique reflections out to 2.2 Å with 73% completeness of the data. The *R* factor for the G3P complex is 13.7% with essentially the same completeness out to the same resolution.

The structures of the protein in the complexes with 3PP, 3PGA and G3P are all very similar despite the differences between the three ligands. The phosphor positions of 3PGA and G3P are of special interest in view of the difference of 1 Å described before between the sulphate position of SO_4^{2-} and the phosphate position of HPO_4^{2-} and 3PP. It turns out that the phosphate positions of 3PGA, G3P, 3PP and HPO_4^{2-} all fall within 0.2 Å of each other. In other words, there seems to be a rather well defined binding site for phosphate- and phosphonate-moieties. This 'phosphor site' is independent of the precise nature of the inhibitor moiety bound to the active site.

In the TIM:3PP complex the side chain of Glu 467 is engaged in a hydrogen bond with the carboxylate groups of 3PP as described above. This is the so-called 'swung-in' conformation of Glu 467. In spite of the presence of equivalent carboxylate groups in 3PP and 3PGA, in the TIM:3PGA complex Glu 467 is in a 'swung-out' conformation. This conformation differs from the 'swung-in' conformation by a rotation of 120° about the $\text{C}^\alpha\text{--C}^\beta$ bond, of 40° about $\text{C}^\beta\text{--C}^\gamma$ and of 60° about $\text{C}^\gamma\text{--C}^\delta$.

In view of the small differences between 3PGA and G3P one might have expected that Glu 467 would adopt similar conformations upon binding these two analogues. This is not the case: the inhibitor G3P binds with Glu 467 in the 'swung-out' conformation.

As seen before, subtle differences in the molecules bound do lead to distinct variations in the mode of binding.

Molecular Modelling of Active Site Inhibitors

The exploration of the binding properties of the active site of gTIM described above are just the very first steps of our 'long' path towards the design of selective inhibitors of trypanosomal TIM. As detailed in table 5, a distance of 12–19 Å has to be spanned from the phosphor atom at the entrance of the active site towards a hydrophobic surface area where significant differences in amino acids between human and trypanosome TIM occur.

Upon careful analysis this 'pathway' appears to consist of a complex mixture of hydrogen-bond donors and acceptors, charged groups, and small hydrophobic areas. One of us (CLMJV), in collaboration with Dr. K. Müller of Hoffmann-La Roche in Basle, has proposed that T-shaped molecules might be suitable for arriving at selective inhibitors. One such T-shaped molecule is shown in figure 12. This molecule can be considered as consisting of four fragments each of which has a different function and is engaged in different interactions with the enzyme:

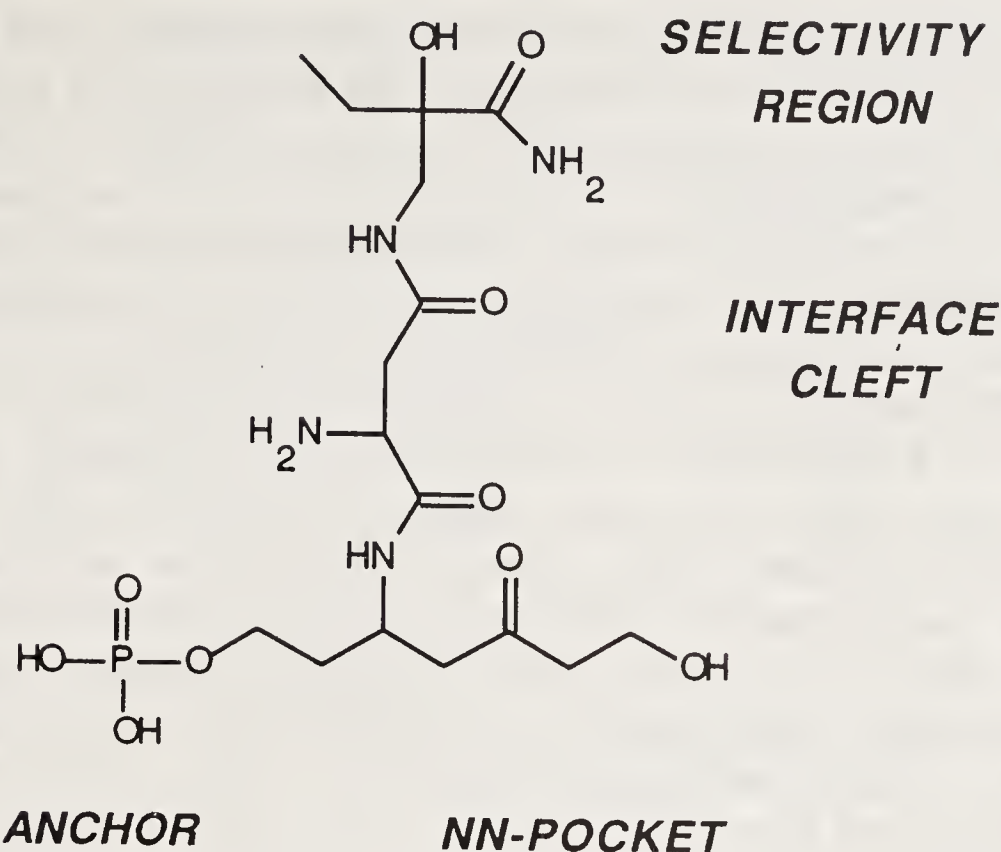


Figure 12. One of the possible 'T-shaped' molecules which might block the entrance to the active site pocket by its phosphate moiety and extend to the region near residues Ala 400 and Tyr 400 of *T. brucei* triosephosphate isomerase.

- (i) The 'anchor' which would block entrance of the active site;
- (ii) The 'NN pocket filler' which extends into a well-defined pocket, made up by Asn 311 and Asn 315;
- (iii) The 'interface' fragment is located at the interface of the two subunits of TIM and links the base of the "T" formed by the two previous fragments with the next one;
- (iv) The 'selectivity fragment' which fits into the hydrophobic area made up by Ala 400 and Tyr 401 and which should provide the inhibitor with the required selectivity versus the human enzyme.

We wish to emphasize that the T-shaped molecule depicted in figure 12 is simply a 'guide molecule' outlining the general dimensions, shape and functional groups. Shortcomings of the modelling process, intervening water molecules, experimental errors and protein adaptability make it virtually impossible that the 'perfect' T-shaped molecule can be designed in one single step. Instead, a 'linked fragment approach' might be more realistic: different functional fragments could first be studied separately. Once several of these fragments, or analogues thereof, are crystallographically characterized when bound to the protein surface, the next step would be to link these fragments together into larger molecules. Such molecules are likely to bind better than the separate fragments, in particular if they could be made more rigid.

Clearly, there is still a long way ahead. But with large amounts of overexpressed *T. brucei* TIM now available, and data collection times being greatly reduced due to area detectors and powerful computers, we hope to increase our structural data base and to improve thereby the accuracy of our molecular modelling efforts.

Interface Loop Based Modelling Studies

As discussed previously, the triosephosphate isomerase active site contains residues from the "interface loop" of the other subunit in the dimer. This loop is indicated

schematically in figure 8 and shown in stereo in figure 13. The exact function of this loop is unknown but it may be responsible for the fact that TIM is only fully active as a dimer, and not as a monomer (60). This fact provides, in principle, another starting point for design of selective *T. brucei* TIM inhibitors: compounds disrupting the *T. brucei* dimer and not the human TIM dimer would decrease the efficiency of trypanosomal glycolysis, while leaving glycolysis of the host unaffected.

With the design of such compounds in mind, the interface loop in the 2.4 Å structure of *T. brucei* TIM (68) was used as a starting point for the design of cyclic peptide analogues by Dr. Klaus Müller in Basle. During the modelling sessions careful attention was paid to one of the many differences in the *T. brucei* TIM interface compared with that of human TIM, namely residue Cys 14 which is a Met in the human enzyme. The interface loop contains a β -turn which consists of Phe 74–Thr 75–Gly 76–Glu 77. One of the cyclopeptides modelled contained this tetrapeptide and consisted of: cyclo[–Phe–Thr–Gly–Glu–Gly–D-Cys–], where the D-Cys might be able to form a disulfide bridge with Cys 14 of *T. brucei* TIM in “subunit-1”. If this cyclopeptide would form a tight complex with single *T. brucei* TIM subunits, then proper interaction of the two subunits in the TIM dimer is prevented.

Inspired by these modelling studies, Dr. R. Osowski in the group of Prof. Kessler

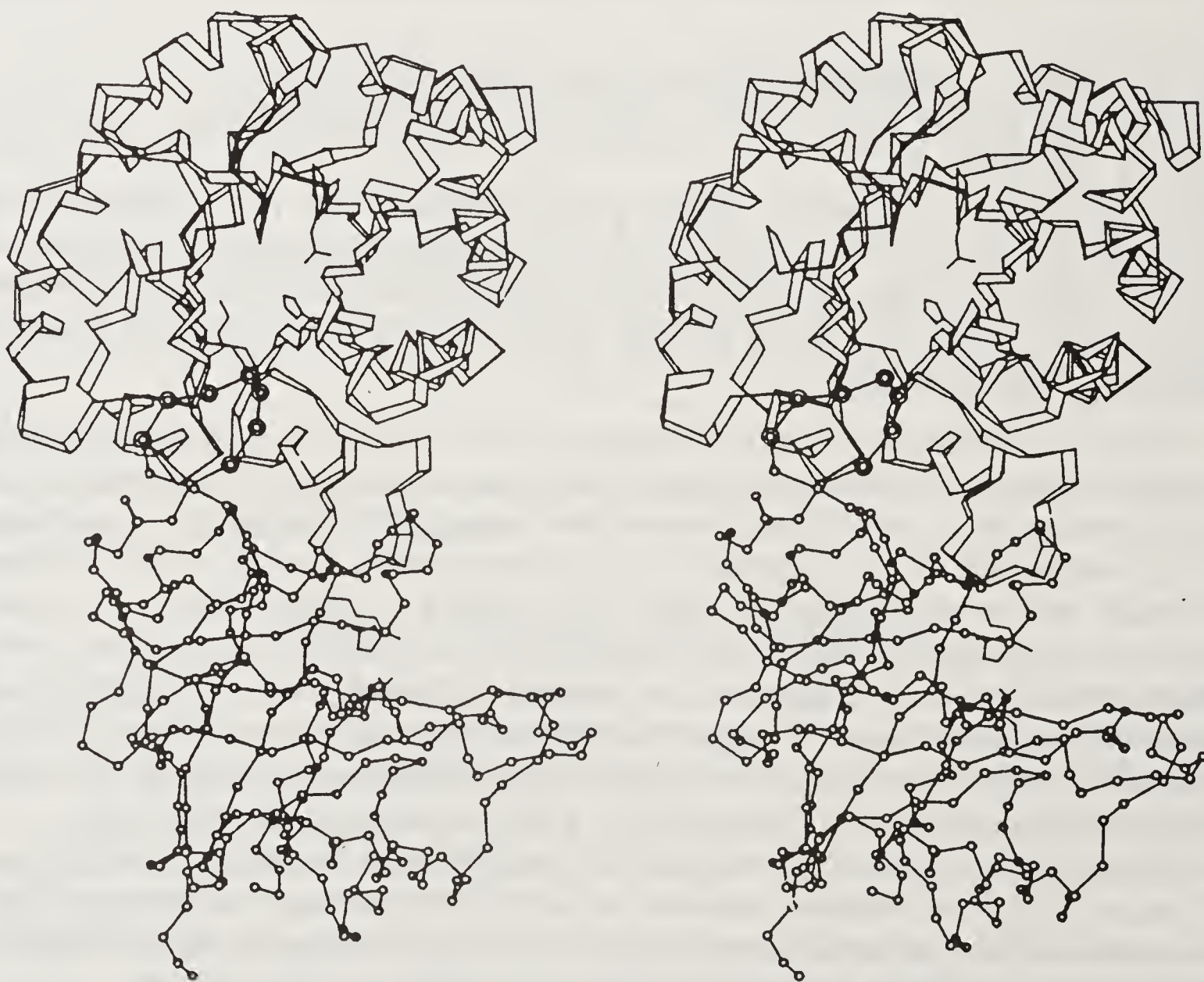


Figure 13. Stereo drawing of the complete gTIM-dimer. The upper subunit (drawn as a continuous ribbon) is “subunit-1”; the bottom subunit (drawn as a continuous line connecting the C α -atoms) “subunit-2”. The active site residues Lys 13, His 95 and Gly 167 of both subunits are indicated. The “interface loop” of subunit-2, indicated by two concentric circles, consists of Gly 72–Ala 73–Phe 74–Thr 75–Gly 76–Glu 77–Val 78.

Table 6. Inhibitory properties of 'serendipitously' discovered cyclopeptide inhibitors of *T. brucei* TIM

Cyclic peptide*	IC ₅₀ -value** (μ M)
c[W-Y(S)-P-F-T(S)-K(Z)]	17.0
c[W-Y-P-F-A-K(Z)]	12.0
c[W-Y-P-F-T-K(Z)]	11.0
c[W-F-P-F-T-K(Z)]	6.5
c[W-F-P-F-F-K(Z)]	2.8

* (Z) = benzene-oxy-carbonyl blocking group

(S) = sulphate group

P is d-proline.

** The IC₅₀-value is the concentration where 50% of the original activity is present.

in Munich, FRG, has synthesized a number of cyclopeptides containing many features of cyclo [–Phe–Thr–Gly–Glu–Gly–D-Cys–]. In enzymatic assays carried out by M. Callens and F. R. Opperdoes in Brussels, it appeared that such peptides had no effect on the enzymatic activity of *T. brucei* TIM. Fortunately, however, Dr. Osowski also took with him a 'reference cyclo peptide' which was just taken from the shelf to serve as a control regarding possible general inhibitory effects of cyclopeptides. By sheer good luck, then, it turned out that this reference peptide – the upper one in table 6 – did have quite a significant inhibitory effect on the activity of *T. brucei* TIM. And in addition it appeared to be a better inhibitor of *T. brucei* TIM than of rabbit TIM. The not entirely illogical supposition that one of the sulphate moieties of the 'reference cyclo peptide' were essential for its inhibitory effects by occupying the phosphate binding site turned out to be wrong. As can be seen in table 6, cyclopeptides with even smaller K_i 's were obtained by removing the sulphate moieties. Eventually, it appeared that the quite hydrophobic cyclo [–Trp–Phe–D-Pro–Phe–Phe–Lys(Z)] had an IC₅₀ of 2.8 μ M for *T. brucei* TIM and an IC₅₀ > 125 μ M for rabbit TIM.

It is clear that we will now try to exploit this stroke of good luck in the continuation of our collaborative project.

Prospects

The above is a description of results obtained by mid-1990 of our project aimed at designing new trypanocidal compounds by focussing on unique features of the energy generating machinery of the sleeping sickness parasite *Trypanosoma brucei*. The first phase of the project has provided a wealth of physicochemical, kinetic, sequence and structural data. This forms a base for the second phase which will concentrate on unravelling details of the import mechanism, elucidating three-dimensional structures of enzymes complexed with a range of ligands, and investigating the effect of newly synthesized compounds on catalysis. At the same time, efforts are continuing to obtain sequence information and three-dimensional structures of additional proteins. With such an expanded structural data base molecular modelling

will become increasingly important to guide the synthetic efforts of the collaborating chemists. We may conclude by expressing the hope that we will eventually not find one, but *several* therapeutically useful compounds each of these interfering with different processes and proteins in the trypanosome, in order to be able to prepare *multi-compound drugs*. Such drugs would suffer much less than single-compound drugs from disastrous drug resistance mechanisms which have been operative in the treatment of other tropical protozoic diseases such as malaria. This is, admittedly, an ambitious goal, but the first steps towards reaching it have been sketched in this overview.

Acknowledgements

It is a pleasure to thank the many scientists who are jointly working on this rational drug design project. In particular we are indebted to Dr. Fred Opperdoes and Dr. P. Michels and their colleagues in Brussels and Prof. P. Borst and co-workers in Amsterdam for a most enjoyable collaboration. Hillie Groendijk has contributed most significantly to the crystallographic studies. We also gratefully acknowledge the enthusiastic support from Dr. Klaus Müller in the modelling efforts. We thank Dr. Jos van Beeumen for gasphase sequenator expertise, Prof. Nickle from Bonn for providing ANTS, and Professors Kessler and Périé for joining with their organic synthetic know-how. We were deeply shocked by the sudden and tragic death of Professor Alan Horn on the second of January 1990. Professor Horn played an important role in supervising the design and synthesis of a variety of active site inhibitors of trypanosomal TIM in collaboration with Cees Witmans and Cor Grol.

The research described was financially supported by the WHO/UNDP/WHO Special Programme for Research and Training of Tropical Diseases, the Alberta Medical Heritage Foundation (RJR); a fellowship from Hoffmann-La Roche (CV); the Programme for Developing Countries of the EC (FV); and the Dutch Foundation for Chemical Research (SON) with financial aid from the Dutch Organisation for Scientific Research (NWO).

We are indebted to Mrs. R. A. Hogenkamp for excellent assistance in preparing the manuscript.

References

1. Perutz, M. F. (1981) in *Atlas of Molecular Structures in Biology*, eds. Phillips, D. C. & Richards, F. M. (Clarendon Press, Oxford), pp. 71–104.
2. Tabin, C. J., Bradley, S. M., Borgmann, C. I., Weinberg, R. A., Papageorge, A. G., Scolnick, E. M., Dahr, R., Lowy, D. R. & Chang, E. H. (1982) *Nature (London)* **300**, 143–149 & Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. (1982) *Nature (London)* **300**, 149–152.
3. Albert, A. (1985) *Selective Toxicity*, (University Press, Cambridge).
4. Hol, W. G. J. (1986) *Angew. Chem., Int. Ed. Engl.* **25**, 767–778.
5. Molyneux, D. H. & Ashford, R. W. (1983) *The Biology of Trypanosoma and Leishmania*, (Taylor and Francis, London).
6. WHO, TDR Seventh Program Report (1985) *Tropical Disease Research*, (WHO, Geneva).
7. Gutteridge, W. E. (1985) *Br. Med. Bull.* **41**, 162–168.
8. Opperdoes, F. R. (1985) *Br. Med. Bull.* **41**, 130–136.

9. Benne, R., Van den Burg, J., Brakenhoff, J. P. J., Sloof, P., Van Boom, J. H. & Tromp, M. C. (1986) *Cell* **46**, 819–826.
10. Shaw, J. M., Feagin, J. E., Stuart, K. S., & Simpson, L. (1988) *Cell* **53**, 401–411.
11. Feagin, J. E., Abraham, J. M. & Stuart, K. S. (1988) *Cell* **53**, 413–422.
12. Simpson, L. & Shaw, J. (1989) *Cell* **57**, 355–366.
13. Weiner, A. M. & Maizels, N. (1990) *Cell* **61**, 917–920.
14. Fairlamb, A. H., Blackburn, P., Ulrich, P., Chait, B. T. & Cerami, A. (1985) *Science* **227**, 1485–1487.
15. Shames, S. L., Fairlamb, A. H., Cerami, A. & Walsh, C. T. (1986) *Biochemistry* **25**, 3519–3526.
16. Krauth-Siegel, R. L., Enders, B., Henderson, G. B., Fairlamb, A. H. & Schirmer, R. H. (1987) *Eur. J. Biochem.* **164**, 123–128.
17. Bitonti, A. J., Bacchi, C. J., McCann, P. P. & Sjoerdsma, A. (1985) *Biochem. Pharmacol.* **34**, 1773–1777.
18. Bacchi, C. J. (1981) *J. Protozool.* **28**, 20–27.
19. Sjoerdsma, A. & Schechter, P. J. (1984) *Clin. Pharmacol. Ther.* **35**, 287–300.
20. WHO, TDR Eight Program Report (1987) *Tropical Disease Research*, (WHO, Geneva).
21. Vickerman, K. (1978) *Nature (London)* **273**, 613–617.
22. Borst, P. (1986) *Annu. Rev. Biochem.* **55**, 701–732.
23. Esser, K. M. & Schoenbehler, M. J. (1985) *Science* **229**, 190–193.
24. Baltz, T., Giroud, C., Baltz, D., Roth, C., Raiband, A. & Eisen, H. (1986) *Science* **319**, 602–606.
25. Pays, E. & Steinert, M. (1988) *Annu. Rev. Genet.* **22**, 107–126.
26. Freymann, D. M., Metcalf, P., Turner, M. & Wiley, D. C. (1984) *Nature (London)* **311**, 167–169.
27. Metcalf, P., Blum, M., Freymann, D., Turner, M. & Wiley, D. C. (1987) *Nature (London)* **325**, 84–86.
28. Wang, C. C. (1984) *J. Med. Chem.* **27**, 1–9.
29. Misset, O., Bos, O. J. M. & Opperdoes, F. R. (1986) *Eur. J. Biochem.* **157**, 441–453.
30. Opperdoes, F. R. (1983) in *Mechanism of Drug Action*, eds. Singer, T. P., Mansoor, T. E. & Undarza, R. N. (Academic Press, New York), pp. 121–131.
31. Fairlamb, A. H., Opperdoes, F. R. & Borst, P. (1977) *Nature (London)* **265**, 270–271.
32. Clarkson, A. B. & Brohn, F. H. (1976) *Science* **194**, 204–206.
33. Opperdoes, F. R. & Borst, P. (1977) *FEBS Lett.* **80**, 360–364.
34. Opperdoes, F. R. (1987) *Annu. Rev. Microbiol.* **41**, 127–151.
35. Opperdoes, F. R., Baudhuin, P., Coppens, I. de Roe, C., Edwards, S. W., Weijers, P. J. & Misset, O. (1984) *J. Cell. Biol.* **98**, 1178–1184.
36. Fairlamb, A. H. & Opperdoes, F. R. (1986) in *Carbohydrate Metabolism in Cultured Cells*, ed. Morgan, M. J. (Plenum Publishing Corporation, New York), pp. 183–224.
37. Misset, O. & Opperdoes, F. R. (1987) *Eur. J. Biochem.* **162**, 493–500.
38. Lambeir, A.-M., Opperdoes, F. R. & Wierenga, R. K. (1987) *Eur. J. Biochem.* **168**, 69–74.
39. Swinkels, B. W., Gibson, W. C., Osinga, K. A., Kramer, R., Veeneman, G. H., van Boom, J. H. & Borst, P. (1986) *EMBO J.* **5**, 1291–1298.
40. Osinga, K. A., Swinkels, B. W., Gibson, W. C., Borst, P., Veeneman, G. H., van Boom, J. H., Michels, P. A. M. & Opperdoes, F. R. (1985) *EMBO J.* **4**, 3811–3817.
41. Michels, P. A. M., Poliszczak, A., Osinga, K. A., Misset, O., van Beeumen, J., Wierenga, R. K., Borst, P. & Opperdoes, F. R. (1986) *EMBO J.* **5**, 1049–1056.
42. Lambeir, A.-M., Loiseau, A. M., Kuntz, D. A. & Opperdoes, F. R. (1991) *Eur. J. Biochem.* **198**, 429–435.
43. Clayton, C. E. (1985) *EMBO J.* **4**, 2997–3003.
44. Marchand, M., Poliszczak, A., Gibson, W. C., Wierenga, R. K., Opperdoes, F. R. & Michels, P. A. M. (1988) *Molecular and Biochemical Parasitology* **29**, 65–76.
45. Marchand, M., Kooijstra, U., Wierenga, R. K., Lambeir, A.-M., van Beeumen, J., Opperdoes, F. R. & Michels, P. A. M. (1989) *Eur. J. Biochem.* **184**, 455–464.
46. Hart, D. T., Baudhuin, P., Opperdoes, F. R. & de Duve, C. (1987) *EMBO J.* **6**, 1403–1411.
47. Clayton, C. E. (1987) *J. Cell. Biol.* **105**, 2649–2654.
48. Wierenga, R. K., Swinkels, B. W., Michels, P. A. M., Osinga, K. A., Misset, O., van Beeumen, J., Gibson, W. C., Postma, J. P. M., Borst, P., Opperdoes, F. R. & Hol, W. G. J. (1987) *EMBO J.* **6**, 215–221.
49. Swinkels, B. W., Evers, R. & Borst, P. (1988) *EMBO J.* **7**, 1159–1165.

50. Kendall, G., Wilderspin, A. W., Ashall, F., Miles, M. A. & Kelly, J. M. (1990) *EMBO J.* **9**, 2751–2758.
51. Michels, P. A. M. (1988) *Biology of the Cell* **64**, 157–164.
52. Opperdoes, F. R., Wierenga, R. K., Noble, M. E. M., Hol, W. G. J., Wilson, M., Kuntz, D. A., Callens, M. & Perié, J. (1990) in *Parasites: Molecular Biology, Drug and Vaccine Design*, UCLA Symp. Mol. Cell. Biol., New Ser., **130**, 233–246.
53. Eilers, M. & Schatz, G. (1986) *Nature (London)* **322**, 228–232.
54. Leslie, A. G. W. & Wonacott, A. J. (1984) *J. Mol. Biol.* **178**, 743–772.
55. Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C. & Rossmann, M. G. (1975) *J. Biol. Chem.* **250**, 9137–9162.
56. Skarzynski, T., Moody, P. C. E. & Wonacott, A. J. (1987) *J. Mol. Biol.* **193**, 171–187.
57. Mercer, W. D., Winn, S. I. & Watson, H. C. (1976) *J. Mol. Biol.* **104**, 277–283.
58. Read, R. J., Wierenga, R. K., Groendijk, H., Lambeir, A.-M., Opperdoes, F. R. & Hol, W. G. J. (1987) *J. Mol. Biol.* **194**, 573–575.
59. Rossmann, M. G. (1972) *The Molecular Replacement Method*, (Gordon & Breach, New York).
60. Waley, S. G. (1973) *Biochem. J.* **135**, 165–172.
61. Knowles, J. R. & Alber, W. J. (1977) *Acc. Chem. Res.* **10**, 105–111.
62. Maquat, L. E., Chilcote, R. & Ryan, P. M. (1985) *J. Biol. Chem.* **260**, 3748–3753.
63. Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, E. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milnzan, J. D., Offord, R. E., Priddle, J. D. & Waley, S. G. (1975) *Nature (London)* **255**, 609–614.
64. Alber, T., Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Rivers, P. S. & Wilson, I. A. (1981) *Philos. Trans. R. Soc. London, Ser. B.* **293**, 159–171.
65. Lolis, E., Alber, T., Davenport, R. C., Rose, D., Hartman, F. C. & Petsko, G. A. (1990) *Biochemistry* **29**, 6609–6618.
66. Lolis, E. & Petsko, G. A. (1990) *Biochemistry* **29**, 6619–6625.
67. Hol, W. G. J. (1985) *Prog. Biophys. Mol. Biol.* **45**, 149–195.
68. Wierenga, R. K., Kalk, K. H. & Hol, W. G. J. (1987) *J. Mol. Biol.* **198**, 109–121.
69. Joseph, D., Petsko, G. A. & Karplus, M. (1990) *Science* **249**, 1425–1428.
70. Wierenga, R. K., Noble, M. E. M., Postma, J. P. M., Groendijk, H., Kalk, K. H., Hol, W. G. J. & Opperdoes, F. R. (1991) *Proteins* **10**, 33–49.
71. Noble, M. E. M., Wierenga, R. K., Lambeir, A.-M., Opperdoes, F. R., Thunnissen, A.-M. W. H., Kalk, K. H., Groendijk, H. & Hol, W. G. J. (1991) *Proteins* **10**, 50–69.
72. Noble, M. E. M., Verlinde, C. L. M. J., Groendijk, H., Kalk, K. H., Wierenga, R. K. & Hol, W. G. J., *J. Med. Chem.*, in press.
73. Lu, H. S., Miao, P. & Gracy, R. W. (1984) *J. Biol. Chem.* **259**, 11958–11968.
74. Schreuder, H. A., Groendijk, H., van der Laan, J. M. & Wierenga, R. K. (1988) *J. Appl. Crystallogr.* **21**, 426–429.
75. Pompliano, D. L., Peyman, A. & Knowles, J. R. (1990) *Biochemistry* **29**, 3186–3194.
76. Rieder, S. V. & Rose, I. A. (1959) *J. Biol. Chem.* **234**, 1007–1010.
77. Wolfenden, R. (1969) *Nature (London)* **223**, 704–705.
78. Collins, K. D. (1974) *J. Biol. Chem.* **249**, 136–142.
79. Belasco, J. G., Herilhy, J. M. & Knowles, J. R. (1978) *Biochemistry* **17**, 2971–2978.
80. Biesecker, G., Harris, J. I., Thierry, J. C., Walker, J. E. & Wonacott, A. J. (1977) *Nature (London)* **266**, 328–333.
81. Straus, D. & Gilbert, W. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2014–2018.
82. Alber, T. & Kawasaki, G. (1982) *J. Mol. Appl. Gen.* **48**, 419–434.

Protein Interaction, Folding and Energetics Studied by Hydrogen Exchange Labeling

S. Walter Englander, Leland Mayne, Russell E. McKinnie,
Yvonne Paterson and Joan J. Englander

Department of Biochemistry & Biophysics and Department of Microbiology,
University of Pennsylvania, School of Medicine, Philadelphia,
PA 19104-6059, USA

1. Contemporary Problems in Protein Structure and Function

Let us start by naming some of the major contemporary problems in protein chemistry.

- (1) *The protein folding problem*: How does a protein, initially synthesized as a randomly structured polypeptide, manage to fold into its complex 3-dimensional structure?
- (2) *The stabilization problem*: What are the design principles by which networks of bonding interactions cooperate to produce the stable native structure?
- (3) *The problem of protein function*: How do the structures of proteins produce their manifold functions? From an experimental viewpoint, how can one determine the parts of any protein that are invested with some particular function, and how does each part help to transduce ligand binding energy and structural energy into a functionally useful form?

These problems are not new ones. They were largely formulated in the 1950s and 1960s. During that period, exciting advances in the understanding of protein structure, to which Professor G. N. Ramachandran contributed so much, made it seem that all these structure–function problems would soon be solved. Yet here we are, 30 years and over 400 protein structures later, and these problems are still with us. We have learned what proteins look like, but not how they work.

Newly emerging concepts and methodologies have in recent years created a renewed sense of enthusiasm that these long-standing problems may soon be solved. This article deals with the application to these problems of hydrogen exchange labeling approaches, which appear to be able to provide information on structure, structure change, dynamics, and energetics at many positions in protein molecules in solution.

2. Hydrogen Exchange Measurement and Interpretation

A small protein molecule like cytochrome c, at 12,500 Da, has 100 backbone peptide group hydrogens. Hemoglobin, at 65 kDa, has about 600 peptide NH. These

hydrogens, along with many polar side chain OH and NH, are in continual exchange with the hydrogens in water. The various hydrogens in a protein exchange over an extraordinary dynamic range, often more than 10 orders of magnitude wide. The exchangeable hydrogens provide many experimentally accessible probe points which, like tiny beacons distributed throughout every protein molecule, continually send out information, in a non-perturbing way, resolved to the level of individual amino acids, concerning structure and interactions at points all through the protein.

Two major problems have tended to limit the usefulness of the hydrogen exchange capability. These are the problem of measuring the exchange behavior of many individually identifiable hydrogens, and the problem of interpreting hydrogen exchange (HX) measurements in terms of structure and energy. Advances in two-dimensional nuclear magnetic resonance (2D NMR) methodology now make it possible to determine the exchange behavior of every individual hydrogen in a small protein molecule like cytochrome c (§3.1). Larger, perhaps more interesting proteins, like hemoglobin, are still beyond the capabilities of NMR methods, but it is possible to measure the exchange behavior of hydrogens in such proteins by use of medium resolution methods (§6.3).

Here we consider some recent experiments that show how protein HX behavior can provide detailed information on interesting protein problems. Along the way, the results obtained raise the problem of HX measurement and interpretation, which we will consider.

3. Protein-Protein Interaction by NMR-Detected HX Labeling

Protein-protein interactions are often characterized by highly specific surface contacts, as in the case of interaction between a protein antigen and an antibody. It has been possible to define a protein's antibody combining region, the epitope, by X-ray diffraction methods in only a few cases. At an amino acid level, protein epitopes may be mapped by hydrogen exchange methods. In recent work, we used the H-D exchange reaction, monitored by 2D NMR, to study the interaction between horse cytochrome c and a monoclonal antibody (1).

3.1 HX Measurement by 2D NMR

Figure 1 shows the 2D COSY spectrum of reduced equine cytochrome c. The diagonal of the 2D spectrum consists of the familiar 1D NMR spectrum. In the contour presentation of figure 1, one is looking down on the spectrum from above, just as in a contour map of the earth's surface. The off-diagonal cross peaks in the COSY spectrum are produced by the through-bond coupling of pairs of near-neighbor protons. For example, the boxed region in figure 1 outlines the so-called fingerprint region which contains cross peaks between each peptide NH and its adjacent alpha carbon CH. We have recently succeeded in assigning each of these cross peaks to their parent protons in cytochrome c, both in the reduced (2) and the oxidized (3) forms (see also 4).

If cytochrome c is placed into D₂O, all the peptide NH in the protein begin to exchange to ND, each at its own characteristic rate. A series of COSY spectra recorded as a function of H-D exchange time then display progressively decreasing amplitudes for the different cross peaks in the fingerprint region, since the ND form is invisible

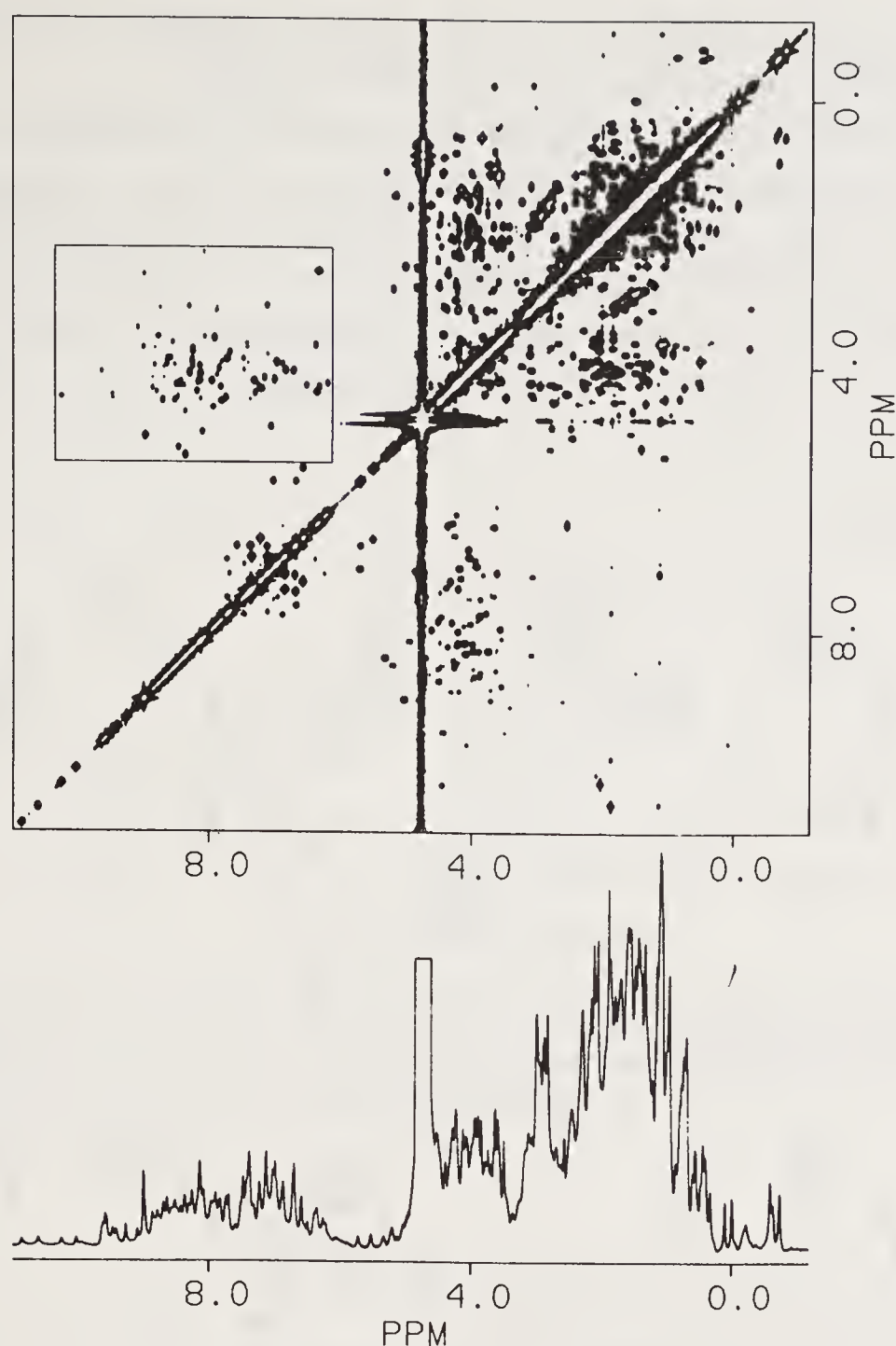


Figure 1. The 1D NMR spectrum of reduced equine cytochrome c measured at 500 MHz (bottom) and the corresponding 2D *J*-correlated or COSY spectrum (top). The box encloses the fingerprint region which contains cross peaks produced by the through-bond coupling of each peptide NH and its neighboring C α H proton.

to proton NMR. From this data, the rate of exchange of each peptide NH proton can be determined.

In the antibody experiment, cytochrome c was bound to a specific monoclonal antibody (named E8) that had been immobilized on an affinity column. The H–D exchange reaction was initiated by washing D₂O into the column. After various H–D exchange times, from 1 hour to 11 days (pD 7, 20 °C), the complex was dissociated by washing the column with buffer solution under somewhat acid conditions (pH 2.7, 8 °C, D₂O), where the H-exchange reaction is very slow. The recovered cytochrome c was then reconcentrated and submitted to 2D NMR analysis in order to determine the level of remaining, unexchanged NH at each amino acid position. NMR spectra were recorded at pD 5.3, where the H-exchange reaction is much slower than for the pD 7 condition of the exchange experiment, so that further H–D exchange is essentially halted. Thus, even though the NMR spectrum requires hours to run, the time resolution

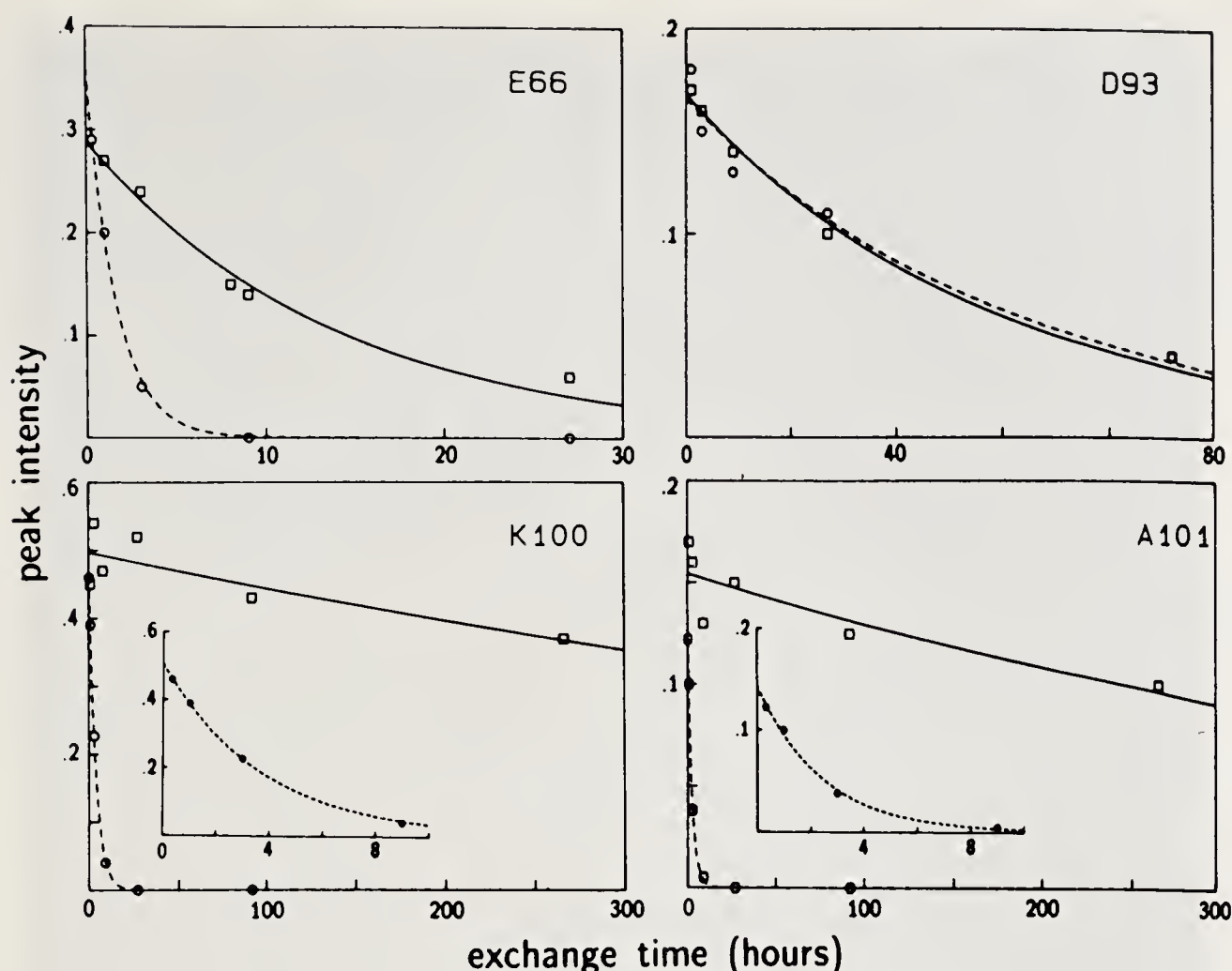


Figure 3. Kinetic curves showing the exchange behavior of some individual peptide NH for cytochrome c either free in solution (dashed curve) or bound to the E8 antibody (solid curve). (reprinted from ref. 1.)

residues in free cyt c, 10, 32, and 95 to 98, still exhibit measurable cross peaks. The others have been exchanged to ND. But in the antibody-complexed samples, an additional 10 NH cross peaks have not yet fully exchanged. H-D exchange kinetics measured in this way for a number of peptide NH is shown in figure 3. In the antibody complex, most cyt c protons are unaffected, but some are slowed by factors up to 340-fold.

3.2 Local HX Effects: Definition of a Protein Epitope

The results identify 11 residues that experience greatly slowed H-exchange in the complex. The protected residues and their H-bond acceptors occur in 3 discontinuous segments of the cyt c polypeptide sequence, but these are all brought together in the folded protein structure to form a continuous largely exposed protein surface, which apparently represents the epitopic region on cyt c for the E8 monoclonal antibody. The surface traced out, shown in figure 4, contains 4 residues that have been identified in previous epitope mapping work performed by chemical protection and mutation experiments (5), and the surface area defined is very similar to other X-ray defined epitopes (6, 7).

3.3 Remote Effects in XH Behavior

In work with another antibody to cyt c, the E3 monoclonal antibody, a totally different epitopic surface is defined (8). The E3 interaction region has less regular

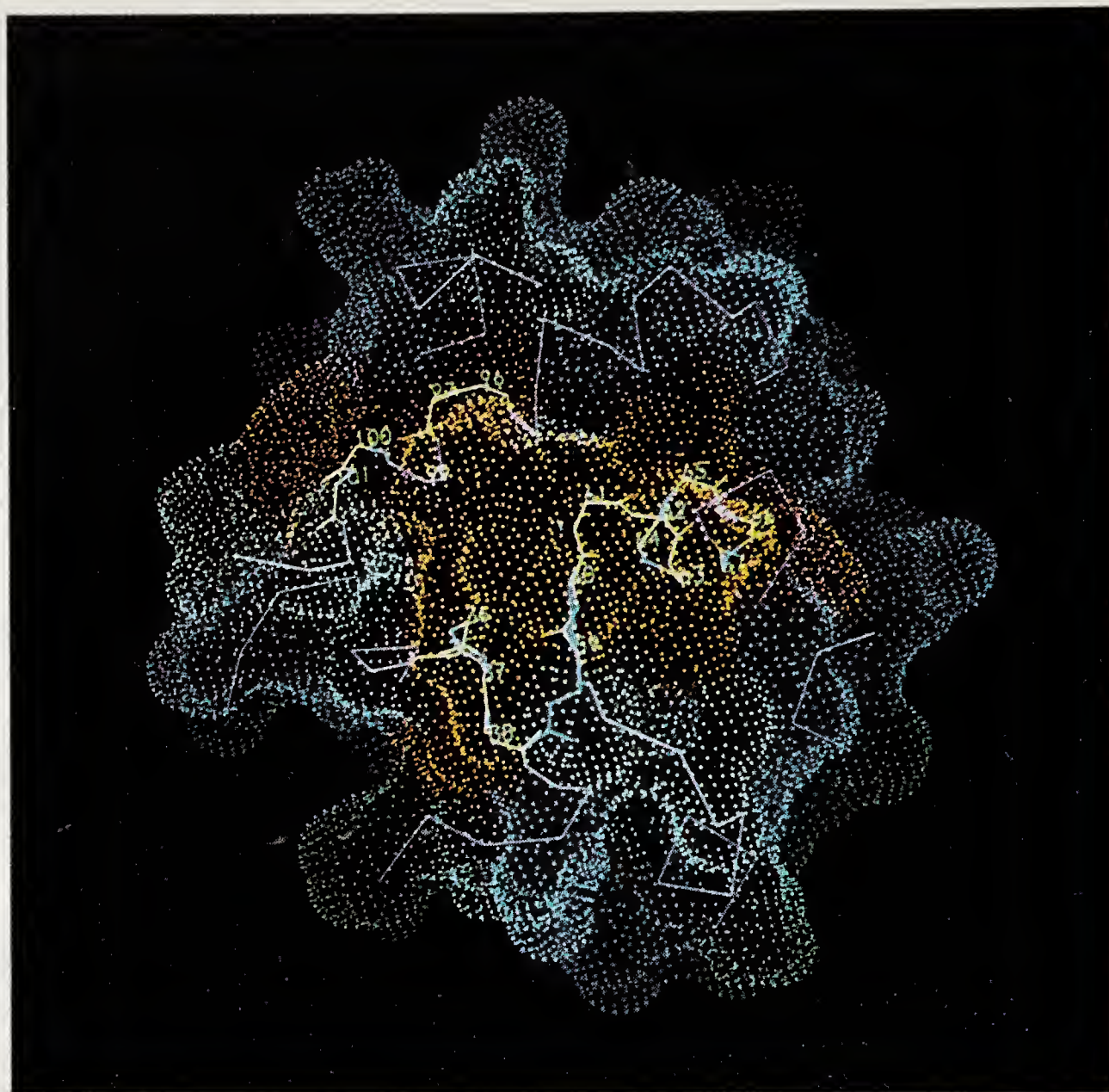


Figure 4. A dot surface representation of the E8 epitope on cytochrome c. The residues identified to have slowed hydrogen exchange and their H-bond acceptors are shown in orange (plus residue 99 identified previously). Residues thought not to be a part of the epitope are colored blue; those not yet identified are in green. (Reprinted from ref 1.)

H-bonding than the E8 epitope, and most interestingly, a remarkable, less simple pattern of HX slowing is observed. The pattern of affected NH makes it clear that the binding reaction can affect the exchange of peptide NH at some distance from the surface actually covered. This kind of behavior is predicted by the so-called local unfolding model for the HX process, which leads one to expect that constraining the protein at one point may produce HX slowing over some distance along the polypeptide chain. The hope of finding and being able to study such cooperative behavior was one of the motivations for performing the antibody experiments initially. Structural models for the HX process, including local unfolding, will be considered below (§ 5).

3.4 Summary

Cytochrome c epitopes for monoclonal antibodies can be defined at an amino acid level by hydrogen exchange labeling. The definition of protein interaction sites represents a straightforward application of the HX labeling approach, and might be extended to nucleic acid and ligand interactions more generally.

4. HX Labeling in the Time Domain: Protein Folding

The capability just described for selective HX labeling of parts of a protein molecule that participate in some particular function can be extended into the time domain. As an example, kinetic intermediates in protein processes can be site-specifically labeled by virtue of differences in the H-exchange behavior of sites affected in the kinetic process. The sites of interest can then be determined by analysis for the H–D exchange pattern in the protein at rest.

4.1 *The Protein Folding Problem*

When a polypeptide chain is released from a ribosome into the potentially hostile biological world, it “knows” how to fold rapidly into its predestined native structure. Even a small protein like cytochrome c, with only 104 amino acid residues, can in principle take up a vast number of different 3-dimensional structures (9). Yet many proteins can accomplish the requisite conformational search and assume their correct structures in a matter of seconds. This process has been demonstrated with a number of pure proteins in the test tube (10, 11). In order to accomplish, within a brief time period, a search that presents an astronomical number of potential intermediates and end points, it seems necessary that the search process be directed through some severely limited number of steps. One therefore suspects that the amino acid sequence of a refolding polypeptide chain must direct it through some more or less well-defined folding pathway.

The difficult problem of discerning the principles by which this extraordinary process occurs has fascinated scientists for decades, since Anfinsen first convincingly demonstrated the independent and spontaneous nature of protein folding (10). The advent of genetic engineering and protein design has greatly expanded current interest in the folding process. One ultimately comes to understand biochemical pathways by recognizing the intermediate structures that define them. Similarly, in order to understand the principles of protein folding pathways, it seems necessary to determine the intermediate structures in some of them. But how is this possible? The intermediate forms in protein folding pathways exist only transiently, do not reach high concentrations, and cannot be isolated for study since they are, by definition, unstable (except for the special case of disulfide crosslinked proteins) (12). Recent work has shown that detailed information on the structure of folding intermediates can be obtained by hydrogen exchange labeling methods.

4.2 *Protein Folding Detected by Spectroscopic Methods*

Figure 5 shows some results for the kinetic refolding of cytochrome c detected by fluorescence methods (13). In these experiments, cyt c denatured in guanidine was diluted into guanidine-free buffer to initiate refolding. The progress of folding was monitored in terms of the decreasing fluorescence of the single tryptophan, Trp 59, which is quenched in native cyt c by resonance energy transfer to the adjacent heme group.

The fluorescence results (figure 5) define the general time scale for the refolding process and manage to resolve several distinct kinetic events. However, this kind of data cannot illuminate the nature of these events. For example, it is possible that

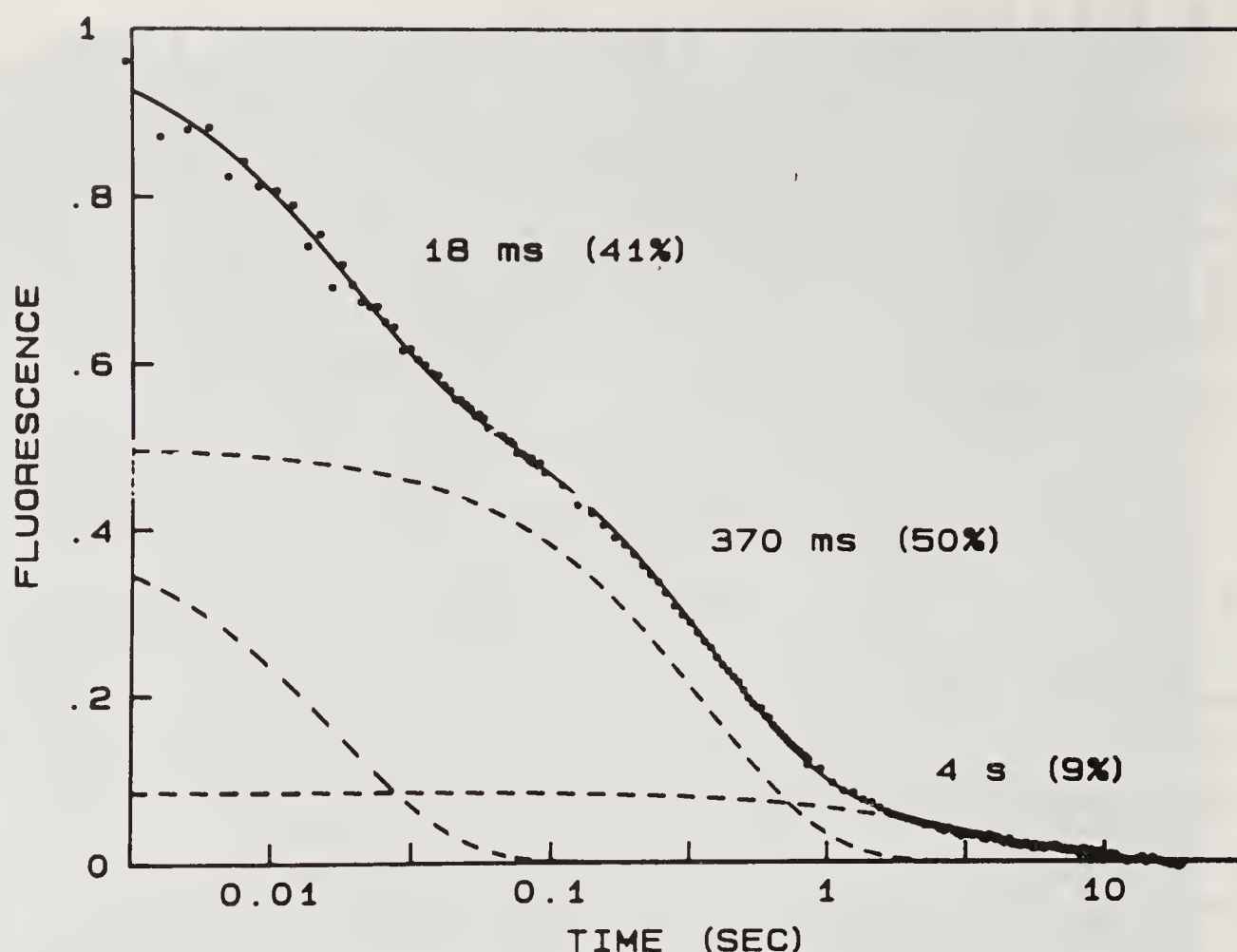
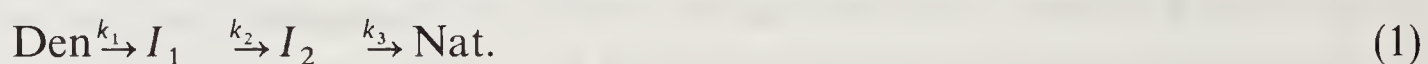
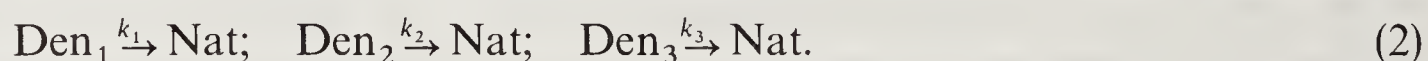


Figure 5. The refolding of cytochrome c monitored by fluorescence emission. The kinetic data are fit by 3 exponentials, shown separately as dashed lines. (Reprinted from ref. 13.)

these kinetics represent several sequential folding events in a homogeneous reaction scheme, as in equation (1). Here every refolding molecule experiences the same series of structural intermediates, with the same kinetic parameters.



At the other extreme, the kinetic steps may reflect a heterogeneous process, as in equation (2), in which different fractions of the refolding population manage to make the complete jump from denatured chain to native protein in a single, apparently 2-state event, with each of several fractions managing the jump on a different time scale.



More complex schemes can be imagined. In addition to these gross uncertainties, the 1-parameter spectroscopic experiment also fails to provide any structural information on potential folding intermediates.

4.3 Protein Folding Followed by HX Labeling

Over 10 years ago, R. L. Baldwin and his coworkers suggested that more information on protein folding intermediates might be obtained by following the refolding process with hydrogen–tritium exchange methods (14, 15). This approach takes advantage of the fact that when an H-bonded structure reforms, the NH involved become protected from H-exchange labeling. Early experiments used tritium exchange labeling in a pH competition mode. The experiment was adapted (16) to H–D exchange labeling for analysis by 1D NMR. More recently, several groups have extended the approach

to H-D exchange labeling in a pulse mode, with analysis by 2D NMR (13,17,18). Some results obtained for cyt c refolding are shown in figure 6.

The HX labeling experiment with cytochrome c parallels the fluorescence experiment just described. Cyt c was initially denatured in guanidine, but here in D_2O so that all the exchangeable NH start out as ND. At time zero, the denaturant level was reduced by dilution of the protein solution into H_2O . The protein began to refold. Although the protein was then in H_2O , exchange of ND to NH does not yet occur since solution conditions were set (pH 6 and $10^\circ C$) so that the HX reaction was relatively slow, with an exchange halftime of several seconds for the freely exposed

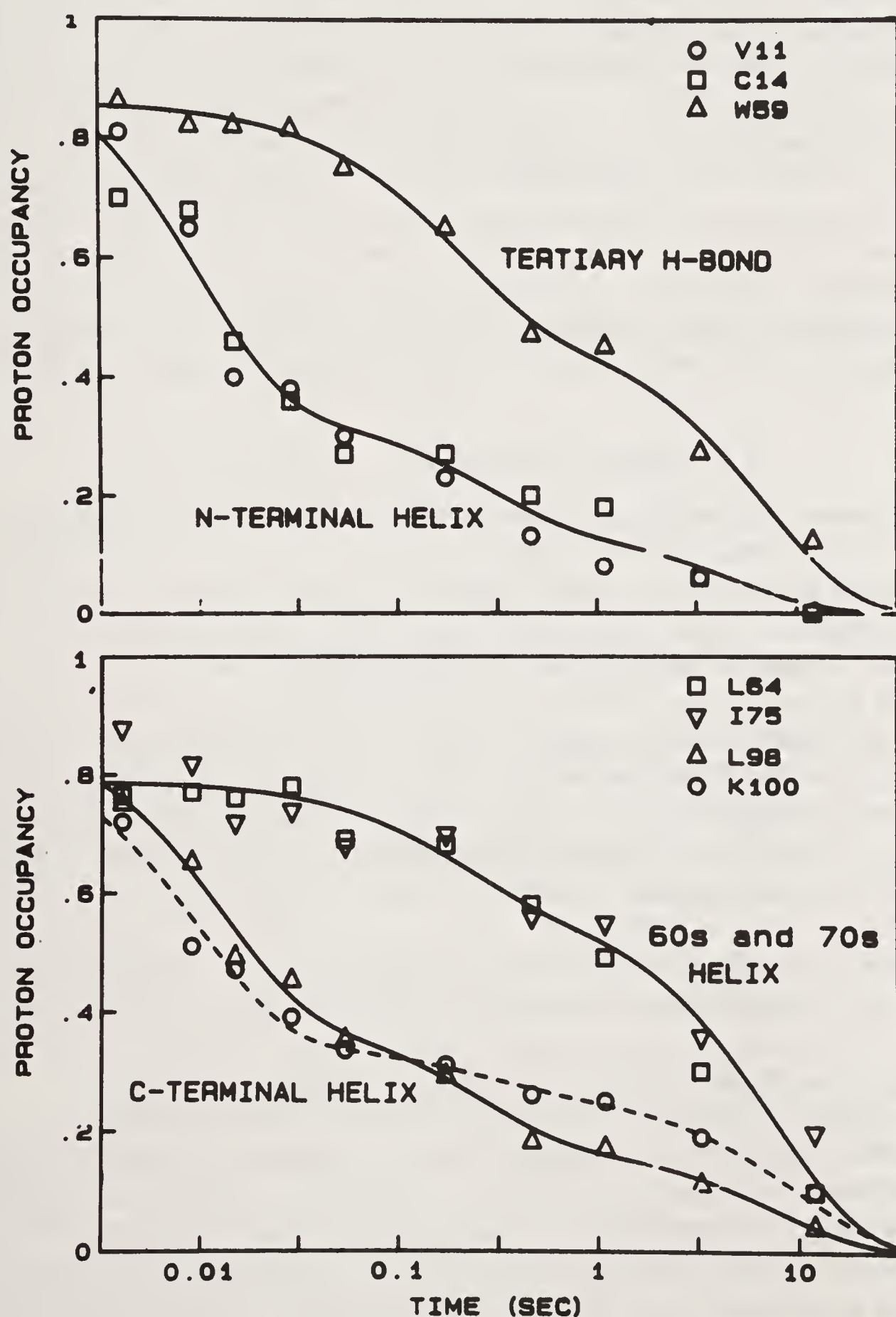


Figure 6. The increasing protection of some peptide NH against HX labeling during the refolding of cytochrome c. (Reprinted from ref. 13.)

peptide group. In order to monitor the structure formed after any time of refolding, the pH of the reaction mixture was pulsed to a higher value (pH 9) to make H-exchange fast (free peptide halftime ~ 2 ms). The pH pulse was maintained for a short time (50 ms), during which period peptide group ND not yet protected by incorporation into H-bonded structure exchange to NH. Sites already involved in reformed H-bonded structure at the time of the pH pulse exchange much more slowly and escape labeling. The labeling pulse was terminated by quenching the pH to a lower value in order to stop further labeling while the protein refolded to its native state.

The stably H-bonded native state together with solution conditions reset to make HX slow (lower pH, reduced cyt c) tend to freeze the H-D exchange labeling pattern inscribed on the protein during the experimental refolding period. A fairly lengthy 2D NMR spectrum of the sample can then be recorded (hours). From the NMR spectrum, the protection against HX labeling achieved by various peptide groups at any time point in the refolding process can be read out by measuring their level of NH labeling. A series of such samples taken at various times during refolding can then portray the time course for formation of individual structural elements of the native protein.

Although experimental manipulations and spectral recording in this analysis consume most of a day, the time resolution of the experiment is shorter by far. This is due to the easy ability to switch the underlying chemical HX rate by many decades simply by changing solution conditions, especially pH. In practice, a time resolution of milliseconds can be achieved, limited only by the rapid mixing apparatus.

4.4 Refolding Intermediates in Cytochrome c

Figure 6 shows the increasing protection against HX labeling experienced by some residues in cyt c during the refolding process (13). Residues in the N-terminal and the C-terminal helices of cyt c achieve protection against labeling, i.e. they generate H-bonded structure, on a time scale of about 20 ms. During this period, residues elsewhere in the protein remain unfolded, and become HX labeled. It is most interesting that, in the native protein, the N- and C-terminal helices are in close contact. The fact that they appear to form simultaneously during folding suggests that the rate limiting step for the formation of both may be their docking to achieve mutual stabilization, as suggested in the diffusion-collision model for protein folding (19). This structure, with the two helices formed and docked while the rest of the protein is still unstructured, appears to define an early folding intermediate in the cyt c pathway.

In addition to structural information, the results in figure 6 contain information on the overall folding process. At time zero, all the sites are unprotected and should exhibit 100% labeling. However, none of the residues are labeled to this level. The data suggest that about 15% of the cyt c molecules refold fully and achieve complete protection against labeling in less than the dead time of the experiment, that is within several milliseconds. The rest of the protein molecules encounter kinetic barriers that place refolding on a longer time scale. About 40% form an initial intermediate, with N- and C-terminal helices folded and apparently docked, on a 20 ms time scale, as just discussed. Another fraction, about 20% of the population, form their terminal helices on a time scale of some hundreds of milliseconds. The remaining 20% do not form their terminal helices until some seconds into the refolding process.

These results show that the refolding protein population is quite heterogeneous.

At least 4 different subpopulations with different folding behavior can be distinguished. For one of these, formation of the terminal helices represents an early step. For the other 3, the terminal helices form later in time and may or may not represent the first step in the refolding sequence of the protein subpopulation involved.

4.5 Summary

The hydrogen exchange labeling approach can be extended into the time domain to obtain overall system-level information and detailed structural information on kinetic protein processes. As exemplified in the protein refolding experiment, the HX labeling approach can combine the spatial resolution available to modern NMR techniques with the millisecond time resolution obtainable by rapid mixing. The experiment works by inscribing structural information on the refolding protein in the form of an H–D exchange labeling pattern that can be frozen and read out later by NMR or other approaches. The results obtained for cyt c refolding demonstrate sequentiality in the folding process and depict qualitatively the structure of an early folding intermediate, but also uncover a perhaps surprising degree of population heterogeneity.

In addition to helping define folding intermediates, the method may be used to measure the stability of the intermediate forms detected, and even of defined regions within each folding intermediate. This may be done by varying the intensity of the HX labeling pulse, using especially pH and time as variables, so that the degree of protection against HX labeling at various positions in any intermediate form can be assessed. The additional capability of hydrogen exchange for measuring local protein stability and its potential significance for protein studies is addressed in the rest of this article.

5. The Protein Hydrogen Exchange Process

Since Kai U. Linderstrom-Lang and his colleagues first embarked on hydrogen exchange studies of proteins in the 1950s, there has been a great deal of uncertainty concerning the mechanism of protein HX processes. The chemistry of proton transfer (20) and of protein H-exchange (21–24) is well understood, but the physical mechanism by which slowly exchanging protein hydrogens are released to solvent is still in doubt. Everyone agrees that observed H-exchange rates depend upon internal protein motions of some kind, but it has been difficult to determine the important details of these dynamical processes, and therefore the proper interpretation of HX data.

5.1 Accessibility-Penetration Models for Protein Hydrogen Exchange

Over the last several decades, two quite different modes for protein HX rate determination have been put forward. A number of accessibility-penetration models (figure 7) have been considered (25–29), based on a common sense point of view. Fast exchanging hydrogens are those accessible at the protein surface; the slow NH are slow because they are buried within the protein. In these models, the effective HX catalysts, OH^- and H^+ ions perhaps along with H_2O , are assumed to penetrate into the protein matrix in order to bring about the HX event.

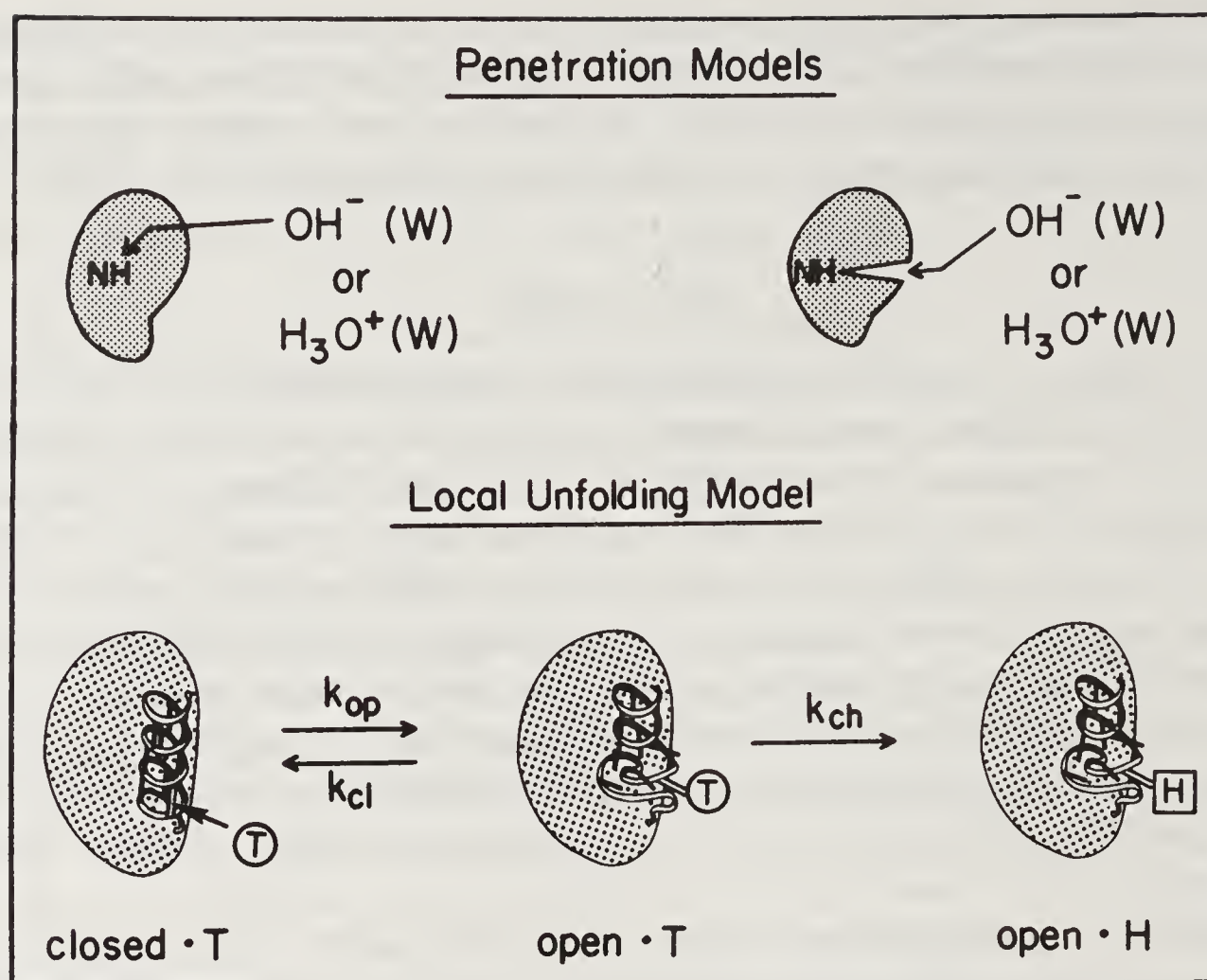


Figure 7. Diagrammatic representations of possible dynamical modes for HX rate determination in proteins.

In our view, this picture predicts the wrong dependence of HX rate on temperature, pH, and viscosity (30). It requires the penetration of charged ions into a hydrophobic milieu, which is energetically most unfavorable (31). In addition there is the surface problem; many of the slowly exchanging hydrogens of proteins are at the aqueous protein surface, which makes it difficult to picture what kind of “penetration process” might be at work.

Observed HX patterns can be taken to suggest an alternative mechanism. We believe that slowly exchanging hydrogens are slow because they are involved in hydrogen bonding, that the exchange process involves the transient breakage of the protecting H-bond, and that due to the cooperative nature of the protein secondary structure, H-bonds tend to break in cooperative sets, referred to as local unfolding units, or unfoldons (figure 7). Local unfolding breaks the protecting H-bonds and brings normally protected NH out into contact with solvent, in a sense the reverse of the penetration picture.

5.2 Unfolding Model for Protein Hydrogen Exchange

The unfolding model, shown in figure 7, draws a deep and potentially most fruitful connection between HX rate and protein structural free energy. The kinetic expression for the limiting case of local unfolding that seems generally to occur indicates that HX rate in a protein (k_{ex}) is lower than the maximum rate characteristic of the freely exposed peptide by the factor, K_{op} , according to equation (3).

$$k_{ex} = K_{op} k_{ch} [\text{OH}^-]. \quad (3)$$

Here $k_{ch}[\text{OH}^-]$ is the HX rate constant characteristic of the solvent-exposed group.

In this picture, K_{op} is the equilibrium constant for breaking the H-bond, or for a local unfolding reaction, essentially the fraction of time the NH is available to the chemical exchange process.

The unfolding model makes some surprising predictions. Sequential NH in residues of an alpha helical segment may be expected to exchange at similar rates, even though some are at the protein surface and others are deeply buried. The NMR capability described above has now uncovered a number of cases with this distinguishing behavior (32–37). Other more complex patterns are also coming to light, in which some NH are considerably slower, or considerably faster, than the group rate. At this time, the correspondence between slow exchange and H-bonding seems secure, but the various kinds of H-bond breakage that may occur and lead to H-exchange, and the fascinating issue of conformation and NH exposure in transiently unfolded states, remain to be fully explained.

5.3 Structural Free Energy

If the unfolding model is roughly correct, then an unprecedented capability becomes available. The measured HX rate of a peptide NH (k_{ex} in equation 3) together with the predictable (23, 24) free peptide rate (k_{ch}) yields the opening equilibrium constant, K_{op} , for the segmental unfolding (or other distortion of structure) that exposes the NH to exchange. In turn, K_{op} measures the structural free energy stabilizing the segment against unfolding, as in equation (4a).

$$\Delta G^0 = -RT \ln K_{op}. \quad (4a)$$

Thus one may be able to define the fundamental cooperative units of protein structure and measure the free energy that stabilizes structure at many points through a protein.

Similarly, one can hope to measure localized changes in free energy associated with structure change, functional state, and interactions. When a change in the HX rate of some hydrogens occurs due to a change in the structural stabilization energy ($\delta\Delta G^0$) of the segment holding them, for example the breakage of a stabilizing salt link as in hemoglobin allostery (below), the relationship connecting these parameters is given by equation (4b).

$$\delta\Delta G = -RT \delta \ln K_{op} = -RT \ln(k_{ex,1}/k_{ex,2}). \quad (4b)$$

Equation (4b) suggests that the measurement of HX rate at any given position in the protein, in the presence ($k_{ex,1}$) and absence ($k_{ex,2}$) of any effect one wishes to study, can reveal the change in structural free energy felt at that position. This treatment is simplified by several assumptions, including the identification of $k_{ch}[\text{OH}^-]$ in equation (3) with the rate characterized for freely exposed small molecule models, but primarily the supposition that the open state visualized in equation (4b) is the same for the two functional states being compared.

5.4 Summary

A variety of models for the protein H-exchange process have been considered over the years. It now appears that H-bonding is a primary determinant of structural slowing, and that H-bond breakage is a necessary event in the HX pathway. A number of examples of locally cooperative H-bond breakage have now been found, although

the resulting HX picture is not always simple. In any case, it appears that this kind of structural determination of HX rate may allow HX measurements to access structural free energy parameters.

The way in which proteins organize their structural energy is at the very basis of their ability to form structure. Further, in order to perform their assigned functions, proteins must in many cases transduce structural energy into functionally useful form. These central parameters, localized structural free energy and free energy change, have not before been generally available to physical measurement. If free energy can in fact be measured as just suggested, then a major capability for unraveling the mysteries of protein structure and function would be at hand. The work with hemoglobin, described below, provides an initial demonstration of this capability.

6. HX Labeling of Structure Change: Hemoglobin Allostery

In order to take up and deliver O₂ effectively, hemoglobin's allosteric mechanism is geared to bind its initial oxygens with reduced affinity and subsequently to adopt an enhanced affinity for all 4 of its ligands. This behavior produces the well-known cooperative binding curve, the signature of the allosteric phenomenon. In terms suggested by equation (4b) and by linked function theory (38), when hemoglobin binds its early ligands with reduced binding energy, the lost binding energy is not destroyed, but is used to strain or actively distort parts of the hemoglobin protein, to raise them to a higher energy level. The allosteric mechanism acts to redistribute this energy through the protein by way of structure change pathways, in a way that causes subsequent binding to occur with higher energy.

We have become used to thinking of structure change as the fundamental touchstone of allosteric process. But equally fundamental is the concept of structural free energy and free energy change. The hemoglobin example is but one illustration of a broad principle. *Structural free energy is the medium of exchange, the negotiable currency, that links ligand binding with structure change and allosteric mechanism.* Structure change plays a material role in that binding energy is transduced by the allosteric mechanism into structure change energy, and free energy is transported through the protein along structure change pathways. In active proteins, structure change may also function to orient particular groups in binding or enzymatic sites, but here too structural free energy must be a central issue.

In order to understand the allosteric process and, more generally, protein function, it will be necessary to measure and localize the fundamental free energy parameter. One has not been used to thinking in these terms because there has been no general way to deal with structure-function problems at this level. Equations (3) and (4) suggest that detailed information on structural stabilization energy may be encoded in hydrogen exchange behavior. But how can the necessary position-resolved and rate-resolved HX measurements be made?

6.1 HX Changes in Large Molecules Measured by Tritium Exchange

Hemoglobin is too large for HX study by contemporary NMR methods. Some alternative analysis is required. Our efforts to localize and quantify HX changes in hemoglobin have progressed through several technological phases. Figure 8 shows

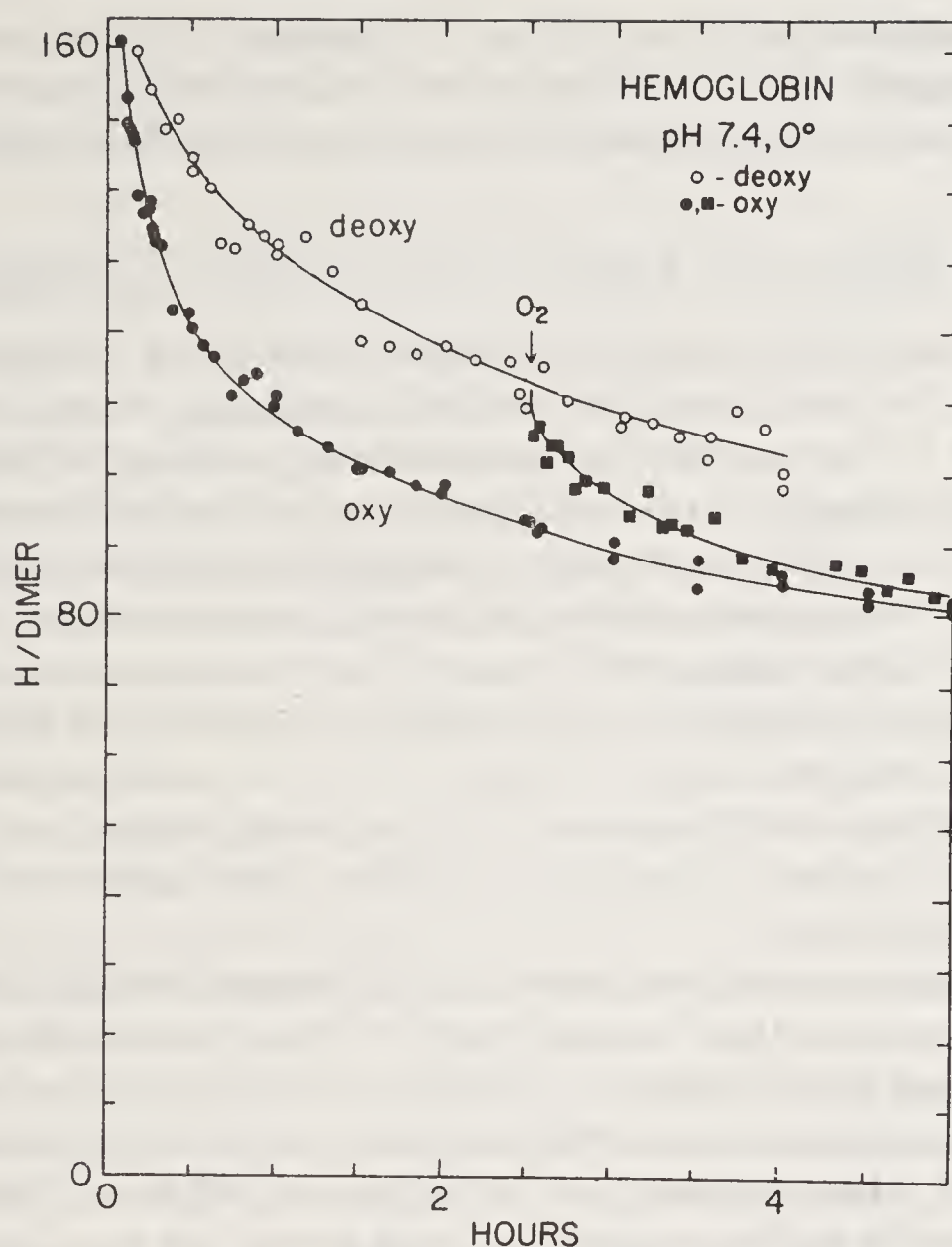


Figure 8. Hydrogen-tritium exchange-out curves for fully labelled hemoglobin in the oxy and deoxy forms, and a crossover curve obtained upon reoxygenation of the deoxy protein.

an early tritium exchange-out curve for oxy and deoxy Hb (36, 39). To obtain this data, oxy Hb samples were initially incubated in tritiated water for an extended time (2 days) under fairly rigorous conditions (pH 9, 37°C), so that even very slowly exchanging sites were labeled. The tritiated protein in either the oxy or deoxy form was then moved into untritiated solvent (pH 7.4, 0°C) by passage through a gel filtration column (~1 min). Subsequent exchange-out of the bound tritium (figure 8) was followed by liquid scintillation counting of protein samples separated from solvent tritium by a second gel filtration.

It has been universally observed that when any protein engages in any interaction or changes its functional state, some changes in its HX behavior occur. Figure 8 shows that hemoglobin is no exception. This kind of data however does not reveal whether all the exchangeable hydrogens in Hb are affected or only some of them, nor what the exchange rates of the affected NH might be in the two protein forms, and certainly gives no information on the identity of the affected protein regions.

The fact that the H-exchange curve of a typical protein is spread over about 10 orders of magnitude on the time axis makes it possible to resolve an HX change in a protein into its component changes in a kinetic sense. The use of tritium exchange labeling then makes it possible to determine the location of the component changes within the protein structure. This can be done by fragmenting the protein and

analyzing the fragments for carried tritium. The selective labeling and the structural analysis that underpin this capability are briefly described in sections 6.2 and 6.3. Then we will return to the fundamental issue of structural free energy.

6.2 Selective HX Labeling at Sites of Functional Change

To selectively label hydrogens at positions involved in change, one can take advantage of the fact that these sites are fast exchanging in one protein form and slow in the other. In the so-called functional labeling method (40), Hb is exposed to tritium exchange labeling for a relatively brief period in the fast exchanging, oxy form. The protein is then switched to the slow exchanging, deoxy form, and exchanged out for a longer time. During the exchange-out period, the tritium label on sites that are unaffected by the change is soon lost, since unaffected sites exchange at the same rate in both protein forms, by definition. But label on sites that are made much slower in the allosteric transition is effectively locked in for a much longer time period. As a result, after some time of exchange-out that is somewhat longer than the exchange-in period, one has a protein sample with tritium label selectively remaining on allosterically sensitive sites.

Figure 9 shows an extreme case. Here hemoglobin was initially labeled in the oxy form for only 1 minute, then was switched to deoxy and exchanged out (all at pH 7.4, 0 °C). Among the NH sites that became labeled during the 1 minute oxy Hb exchange-in, 4 sites (peptide group NH) exchange-out in deoxy Hb on a time scale of several hours; 3 others exchange with a halftime of 100 hours! When a sample of deoxy Hb prepared in this way is switched back to the oxy form, at any time point in figure 9, the still bound tritium label reverts to its fast form and exchanges out in

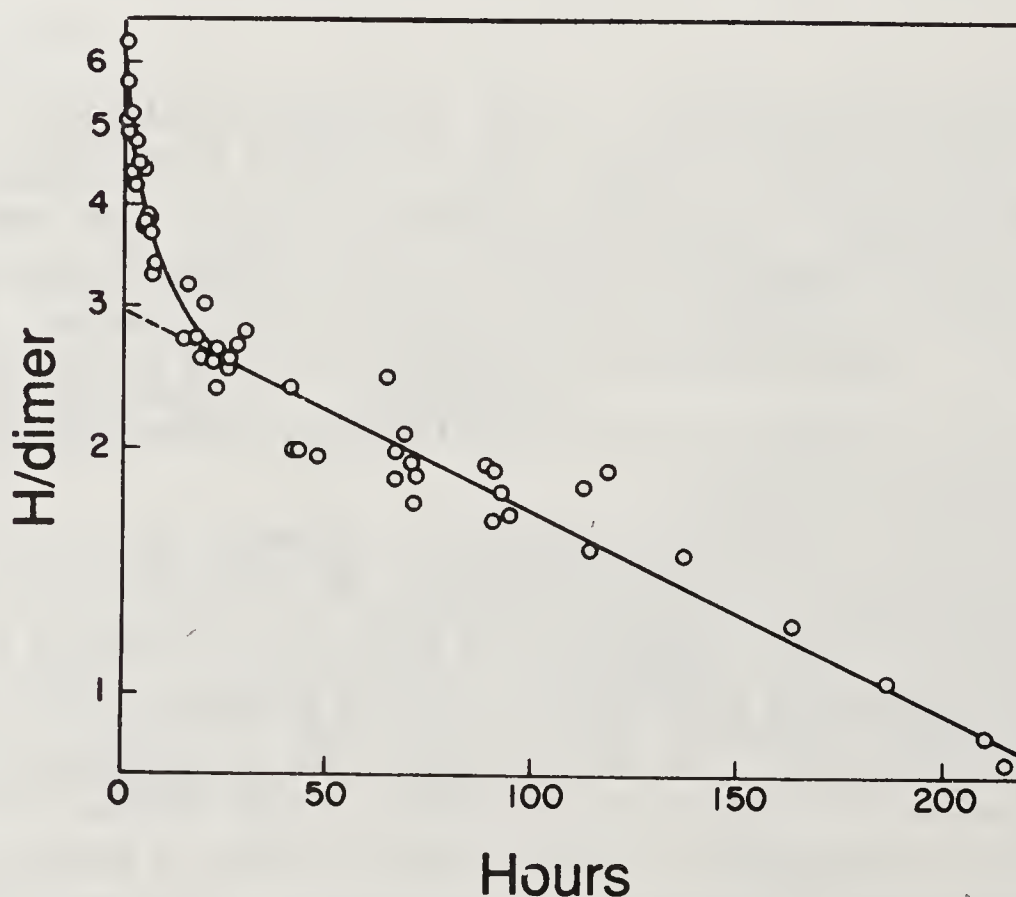


Figure 9. The so-called jump class of allosterically sensitive NH, selectively extracted from the overall HX curve of figure 8 by the functional labeling method. Hemoglobin was initially labeled in the oxy form for 1 minute, then made deoxy and exchanged out.

1 minute. We do not yet know where the 3 extremely slow NH seen in figure 9 are placed in the protein, but we have located the faster phase NH (§ 6.4).

In functional labeling, the timing protocol used for the fast-form exchange-in and the slow-form exchange-out sets up a kind of kinetic filter, a window, that selects certain function-sensitive NH for labeling. In a series of experiments in which the exchange conditions (time, pH, temperature) are varied, one can sweep this window over the multi-decade time scale of the overall HX curve, pick out the NH that are fast in one protein state and slow in another, and measure the number of sensitive NH and their exchange rates in both forms.

Functional labeling does not always work out as cleanly as in figure 9. The ability to selectively study particular NH depends upon the multiplicative difference in their exchange rates in the two protein forms. When these rates are exceedingly different, as in figure 9, the filter works straightaway. When the rates are fairly close, the insensitive NH that exchange within the necessarily overlapping exchange-in and exchange-out time scales are not fully rejected by the kinetic filter, and provide an unwelcome background contribution to the measured HX curve. We have developed methods for selectively portraying this background. In essence, the exchange curve for the background, insensitive NH is obtained by performing the functional labeling procedure in the reverse direction, with initial exchange-in in the slow form. A refined HX curve for the sensitive NH can be obtained by manual subtraction of the background curve. Various examples of these manipulations can be found in references dating back from (41). Equations describing these operations are given in (42).

These methods have been used to survey the entire sweep of the hemoglobin H-exchange curve (see 41, and earlier work). The results show that only a limited fraction of hemoglobin's peptide NH are sensitive to the allosteric transition and that the sensitive NH sites are always faster exchanging, i.e. are destabilized, in the oxy Hb (R state) form. The sensitive NH appear to occur in coherent groupings, 3 to 12 hydrogens in size. The NH in each set exchange within some narrow range of rates in oxy Hb; in the deoxy form the whole group moves in unison to a new rate, slower by 15 to 10^4 -fold. This is just the behavior expected for local unfolding units (figure 7). In fact, it was these results that initially suggested the local unfolding model for protein H-exchange (43).

6.3 *Locating the Functionally Labeled Sites*

We want to identify the functionally involved protein sites that the functional labeling method tags with radioactive tritium. In principle, this can be done by fragmenting the Hb, then separating and analyzing the fragments. This may seem foolish; will not the exchangeable label be lost during the separation process? It has proven possible to accomplish this (44–46) by taking advantage of the fact that the exchange halftime of even fully exposed peptide NH is about 1 hour at pH ~ 3 and 0°C (23). In the fragment separation method (47), samples of Hb selectively labeled as just described are shifted into this slow exchange condition (brief gel filtration), proteolyzed briefly with an acid protease, and the fragments are quickly separated by HPLC under the same slow exchange conditions. During this procedure (~ 30 minutes total time), losses of 20 to 35% of the carried label are experienced, but this can be independently calibrated and corrected for.

Figure 10 exhibits an HPLC fragment separation from a Hb sample in which the

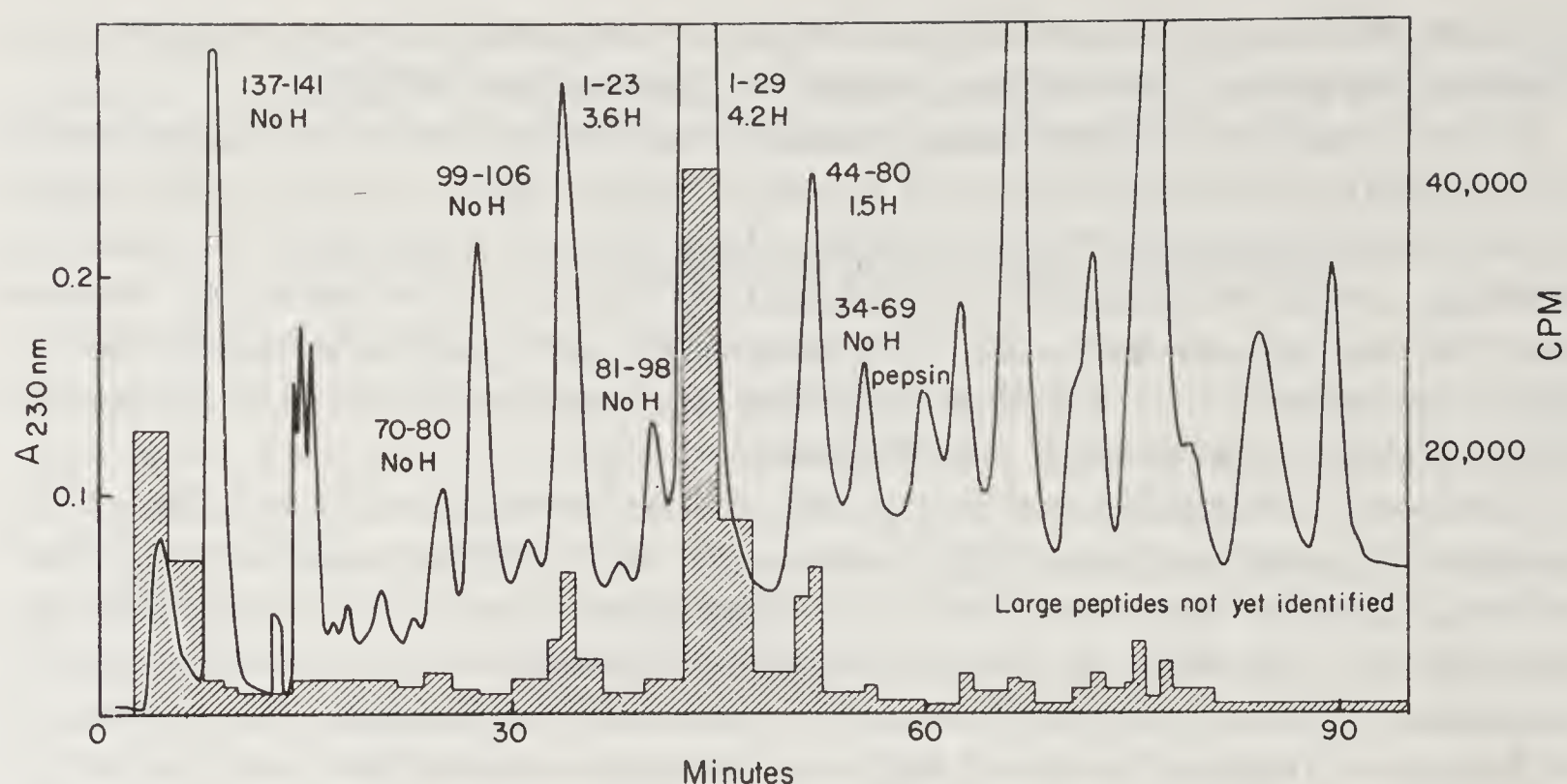


Figure 10. Allosterically sensitive NH near the N-terminus of the hemoglobin alpha chain, isolated from functionally labeled hemoglobin by the fragment separation method. Some of the fragments separated by HPLC are identified, and the number of labeled NH each one carries is shown. The shading indicates the measured tritium count levels.

initial functional labeling protocol placed tritium on 5 allosterically sensitive peptide NH at the alpha chain N-terminus. The demonstration experiment in figure 10 locates these NH on two overlapping fragments, α 1-29 and α 1-23. In a real experiment, the HPLC gradient is tailored to deliver, more specifically and rapidly, just the targeted fragment. Subfragmentation experiments with α 1-29 indicated that the labeled positions are between residues 1 and 12 on the alpha chain (48). Similar work using other labeling protocols has located a set of 7 allosterically sensitive NH within the region β 89-98 on the F-FG segment (placed just behind the beta chain heme and including the proximal His 92 which coordinates the heme iron) (47). Another set of 4 very sensitive NH, isolated on the fragment β 130-146, has been found at the beta chain C-terminus (35, 36).

We assume that allosterically sensitive segments identified in this way mark positions that play a significant role in hemoglobin's allosteric mechanism. In support of this view, the literature provides strong evidence implicating allosteric events at each of the 3 sensitive segments we have so far identified, and all 3 intersect the $\alpha_1\beta_2$ subunit interface, where allosteric function is thought to be concentrated (49). Further, unlike most other measures of structure change (diffraction, spectroscopy, reactivity), wherein one cannot specify the relative importance of an observed change, a change in HX rate is connected by local unfolding theory with change in structural free energy (equation 4), which appears to be the proper measure of relative functional importance.

6.4 Allosteric Free Energy Change at the Beta Chain C-Terminus

The functional labeling-fragment separation approach can be used to measure the exchange behavior of each set of allosterically sensitive NH in the oxy or the deoxy state, or in any experimentally modified form one wants to study. As an example,

figure 11 shows some HX results obtained in this way for the H-bonded peptide NH shown in figure 12 near the C-terminus of the β -chain. These are the NH that account for the faster kinetic phase of the allosterically sensitive HX behavior seen in figure 9. According to crystallographic results (50, 51), the crosslinks indicated by shading in figure 12 exist in deoxy Hb and are severed in the oxy form.

The 4 C-terminal peptide NH indicated in figure 12 exchange with an average halftime about 5 hours in deoxy Hb (figure 11). When Hb is liganded, these NH all move in unison to a new rate 1200-fold faster. This is the hallmark of a cooperative unfolding unit. According to equation (4b), the 1200-fold change in HX rate indicates that the segment is destabilized by a total of 3.8 kcal (per $\alpha\beta$ protomer) in the allosteric transition (pH 7.4, 0 °C, with phosphate). This value may represent the allosterically sensitive free energy that impinges on this segment. If so, then a large fraction of the total allosteric energy in hemoglobin is felt at the immediate beta chain C-terminus.

6.5 Components of the Allosteric Mechanism

We treat the sensitive peptide NH as marker protons. By studying their exchange behavior in the presence of defined chemical and mutational modifications, it may be possible (equation 4b) to measure the localized stabilization free energy that defined linkages provide, individually and in concert.

Louie *et al.* (35, 36) found that when the allosteric effector, DPG, is removed from deoxy Hb, all the marker NH at the beta chain C-terminus exchange 2.5-fold faster,

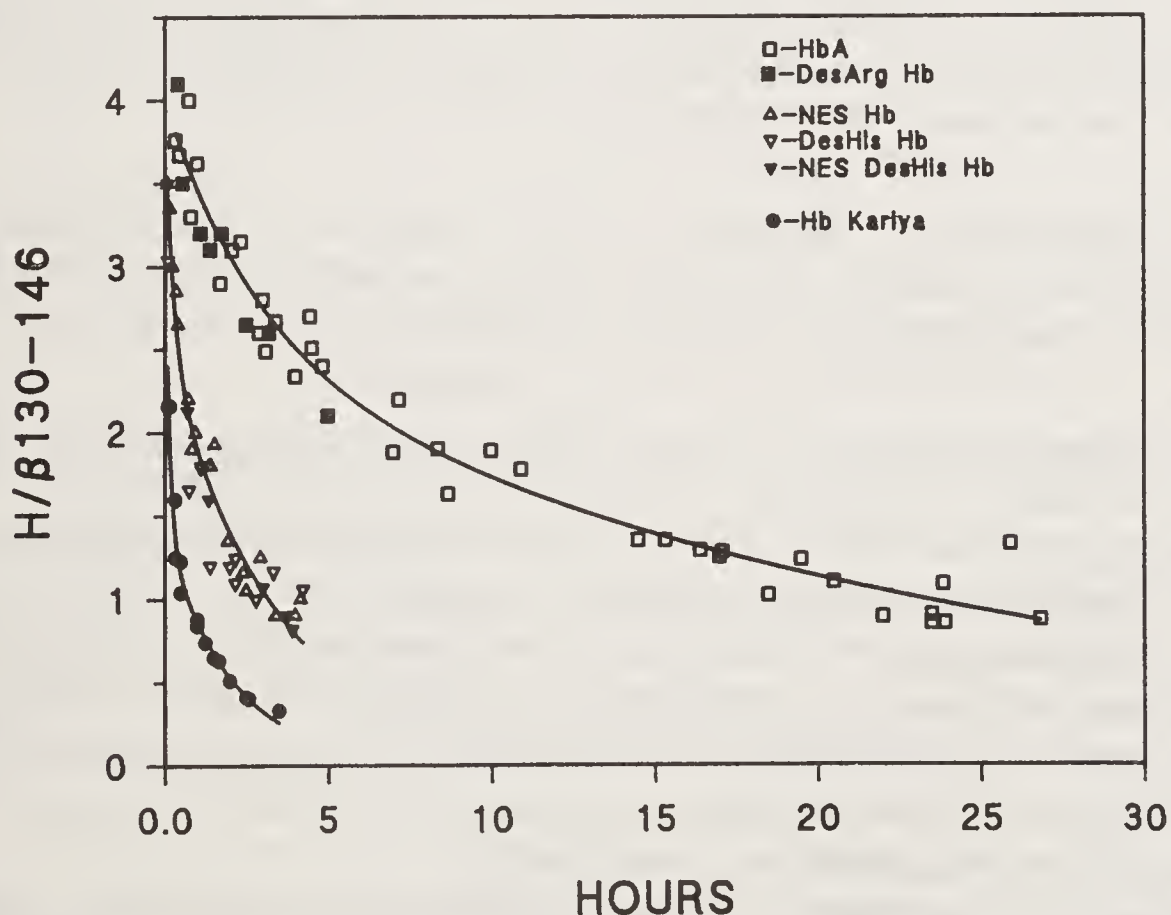


Figure 11. H-exchange behavior of the allosterically sensitive NH at the beta chain C-terminus studied by the functional labeling and fragment separation methods. These NH were isolated and measured on the β 130–146 fragment. The NES modification, or proteolytic removal of the terminal histidine, or both modifications together, cause all the NH to exchange 8-fold faster. In Hb Kariya (Lys 40 α \rightarrow Glu), the marker NH are accelerated by a factor of 36. The remote desArg modification has no effect.

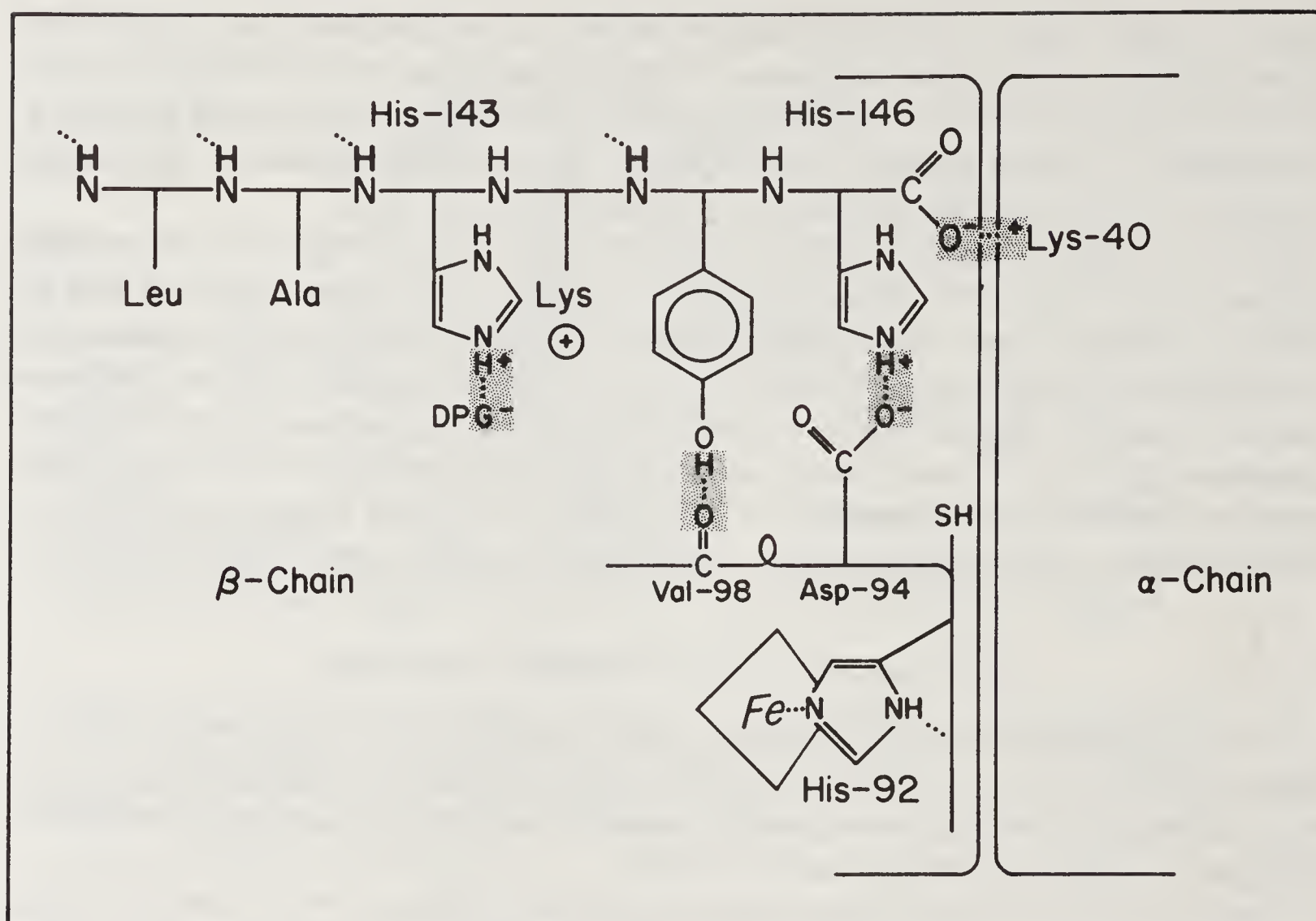


Figure 12. Diagram of some allosterically sensitive linkages (50, 51) near the beta chain C-terminus, and the $\alpha_1\beta_2$ contact region. Groups shown in the beta chain include the heme and its coordinated proximal histidine (His 92), the reactive sulfhydryl (Cys 93), the F-FG segment, and the C-terminal segment. The apparent H-bonded NH on the C-terminus that are measured in figure 11 are shown with dotted lines indicating the H-bonds. (Reprinted from ref. 35.)

indicating a free energy for the DPG to His 143 bond (see figure 12) of 0.5 kcal. This is close to the allosterically effective free energy contribution measured for DPG by other methods. The HX effect measured was sensitive to salt conditions, as has been seen before for the effect of DPG on the O_2 binding curve. Early crystallographic results (50) showed that when the reactive SH (see figure 12) is modified with N-ethyl maleimide, the bulky N-ethyl succinimidyl (NES) derivative breaks one or more bonds to the terminal His 146. HX results (35) show that the NES modification causes all the marker protons to exchange 6.3-fold faster, corresponding to a free energy change of 1.0 kcal. When both the NES modification and removal of DPG were tested together (35, 36) the marker protons all exchanged faster than in unmodified HbA by 16-fold, pointing to a free energy change of 1.5 kcal, equal to the effect expected as the sum of the DPG and NES effects measured separately (1.0 + 0.5 kcal). It appears that these bonds provide simply additive free energy effects.

Which bonds does NES break? Figure 11 shows some more recent results. Under the salt conditions used here, the NES modification causes all the marker protons to exchange 8-fold faster, indicating that the C-terminal segment is destabilized by 1.1 kcal. In desHis146 Hb, the marker protons exchange 8-fold faster, just as in NES-Hb. In desHis146 Hb, the terminal histidine is proteolytically removed, and this destroys both the internal and cross subunit salt links shown in figure 12. This suggests that the NES modification also breaks both the salt links. But again, perhaps the two

different modifications affect different bonds and the free energies just happen to be degenerate. This can be checked by an identity test in which both the NES and desHis modifications are imposed together. If they destabilize different positions, then the two together should have a larger effect than either one separately (as for the -DPG + NES case). Results are shown in figure 11. The double modification has just the same effect as either one separately. We conclude that either modification removes the two histidine salt links, and both modifications together impose no additional effects.

6.6 A Long Range Effect: The Allosteric Network

The results just described show that the free energy calculation (equation 4) performed here exhibits good internal consistency, and suggest that it may be possible in this way to trace out the network of energetic interactions that together comprise the allosteric mechanism. However, the effects described so far are all immediately local to the modifications tested. How does the energetic signal travel to distant hemes?

It is fascinating that some of the modifications just described can be seen to affect distant regions of the protein. As indicated in figure 10, the HX methods detect another allosterically sensitive region at the alpha chain N-terminus, a position that is structurally distant from the beta chain C-terminus. When the desArg141 α modification is imposed (an allosterically involved crosslink connecting the α_1 -chain N-terminus to the α_2 -chain C-terminal arginine is destroyed) no effect is seen by HX at the beta chain C-terminus (figure 11). However, alterations at the beta chain C-terminus can be felt at the alpha chain. The NES and desHis modifications each cause the allosterically sensitive NH at the alpha chain N-terminus to exchange faster by 2-fold, indicating a destabilization of the N-terminal segment by 0.4 kcal.

We have also tested in this way Hb Kariya, in which the cross subunit salt link (figure 12) is destroyed by changing Lys40 α to aspartic acid. This modification, as expected, destabilizes the beta chain C-terminus (by 2 kcal). It also affects the remote alpha chain N-terminus, by the same value, 0.4 kcal, seen for NES and desHis Hb.

6.7 Comparison with Energy Effects Measured by Other Methods

The measurement of the fundamental parameter, allosteric free energy, has been accomplished before in several ways, most notably in subunit dissociation (52) and oxygen binding (53) experiments. It is most interesting to compare these results with the free energy values estimated by the H-exchange method. These comparisons all relate to pH 7.4, but one should note that the HX experiments were done at 0 °C, the subunit work was at 21.5 °C, and the ligand binding curves were measured between 25° and 37 °C. Also, the subunit and liganding experiments measure the global free energy change; the HX measurement is more pointed, and monitors localized effects.

The NES modification, measured both by ligand binding (54) and subunit dissociation (52), is found to destabilize hemoglobin by 1.4 kcal per $\alpha\beta$ protomer. The desHis modification, measured in ligand binding experiments (55), registers 1.3 kcal. The HX results for these modifications point to a 1.1 kcal destabilization at the beta chain C-terminus and 0.4 kcal at the alpha chain N-terminus, for a total of 1.5 kcal, in good agreement with measurements for the global effect.

The Hb Kariya mutation accelerates the beta chain protons (figure 11) 36-fold, indicating 2 kcal of destabilization at the beta C-terminus, and it also affects the alpha chain N-terminus by 0.4 kcal, for a total effect of 2.4 kcal. Similarly, Hb Kariya tested by subunit dissociation (56) and by ligand binding (57) causes a global destabilization of 2.6 kcal per $\alpha\beta$ dimer).

This close agreement is most encouraging. However, it is interesting that not all the effects measured so far find such an easy interpretation. Most striking is the desArg modification, which has a major effect when measured by subunit dissociation (52), destabilizing the tetramer by 2.1 kcal per $\alpha\beta$ protomer. HX registers a local effect at the alpha chain N-terminus, worth only 0.7 kcal, and no effect at the beta C-terminus (figure 11).

Other observations are similarly provocative. Deprotonation of the terminal His146 β (figure 12) appears to account for about half the Bohr effect. The excess free energy (due to the internal histidine salt link?) stabilizing this proton in the deoxy Hb form has been estimated at 1.4 kcal (58–61). However, when we test Hb Cowtown (His146 β \rightarrow Leu, so that the internal histidine salt link is destroyed), only 0.3 kcal in destabilization free energy is registered. This is close to the global value of 0.25 kcal found in subunit dissociation experiments (56). Here is an evident discrepancy. How is the Bohr proton stabilized?

Clearly, the resolution of these localized energy interactions and the understanding of how they are integrated to comprise the global allosteric mechanism will require more work. The optimistic view is that the tools necessary for this work may now be in hand.

6.8 Summary

HX labeling methods that are able to provide site and rate-resolved hydrogen exchange data at positions of allosteric sensitivity are now available. The results obtained so far have significance for HX theory and methodology as well as for understanding Hb allostery.

Cooperative segmental H-exchange behavior has been found for sets of allosterically sensitive NH in hemoglobin. In each segment, the several NH all exchange at about the same rate in deoxy Hb, and when even a single stabilizing linkage to the segment is broken, or when more substantial change occurs as a result of the overall allosteric transition, all the NH in the segment move in unison to a new rate faster by the same factor. In a number of cases where it has been possible to make comparisons, values for change in structural free energy obtained by other methods are in agreement with values found by interpreting HX data according to local unfolding theory. These observations suggest that local unfolding does occur and does dominate HX rates in these particular cases. Some other observations seem anomalous in this respect, and remain to be explained.

In respect to hemoglobin function, results have so far identified 3 centers of allosteric activity, and measured the overall allosteric free energy felt at each. The free energy of individual allosterically important linkages has been measured, and a long range effect has been found. This is just the kind of information necessary to identify the parts of the molecule that are actively involved in the allosteric function and to learn how these parts are networked together to produce the energy-coupled allosteric mechanism.

7. Hydrogen Exchange Labeling and Protein Structure and Function

We started out by listing in general terms some of the major problems in protein chemistry – the folding problem, maintenance of the native state, and the generation of function. To attack these problems it will be necessary to apply methods that can delineate structure, locate the parts of a protein involved in function, and measure structural energy and function-related energy change.

In this article, the potential of hydrogen exchange labeling methods was examined by reviewing HX theory and evolving methodology in the context of several recent applications to significant structure–function problems. The ability of HX labeling to define the interaction region of a protein was considered in a study of a protein–monoclonal antibody interaction. Extension of the HX labeling approach into the time domain was demonstrated in a study of structural intermediates in protein folding. The apparent ability of available HX methods to pinpoint allosterically involved sites and to detail function-related change in structural free energy was shown in a study of allosteric mechanism in hemoglobin.

The HX labeling approach exploits function-based kinetic differences in exchangeable hydrogen sites to selectively label protein positions involved in the function. The elements of the approach, change in HX rate and covalent labeling, seem widely applicable in protein studies. Methods for the measurement of position-resolved hydrogen exchange behavior are now available and advances are being made in the understanding and interpretation of HX mechanism. The applications described here illustrate these methods and advances and show how HX labeling approaches can contribute to the solution of contemporary protein structure–function problems.

References

1. Paterson, Y., Englander, S. W. & Roder, H. (1990) *Science* **249**, 755–759.
2. Wand, A. J., Di Stefano, D. L., Feng, Y., Roder, H. & Englander, S. W. (1989) *Biochemistry* **28**, 186–194.
3. Feng, Y., Roder, H., Englander, S. W., Wand, A. J. & Di Stefano, D. L. (1989) *Biochemistry* **28**, 195–203.
4. Gao, Y., Boyd, J., Williams, R. J. P. & Pielak, G. J. (1990) *Biochemistry* **29**, 6994–7003.
5. Paterson, Y. (1989) *The Immune Response to Structurally Defined Proteins: The Lysozyme Model*, eds. Smith-Gill, S. & Sercarz, E. E. (Adenine Press, Schenectady N.Y.), pp. 177–189.
6. Davies, D. R., Sheriff, S. & Padlan, E. A. (1988) *J. Biol. Chem.* **263**, 10541–10544.
7. Colman, P. M. (1988) *Adv. Immunol.* **43**, 93–132.
8. Mayne, L., Paterson, Y., Cerasoli, D. M. & Englander, S. W. (To be submitted).
9. Dill, K. A. (1985) *Biochemistry* **24**, 1501–1509.
10. Anfinsen, C. B. & Haber, E. (1961) *J. Biol. Chem.* **236**, 1361–1363.
11. Kim, P. S. & Baldwin, R. L. (1990) *Annu. Rev. Biochem.* **59**, 631–660.
12. Creighton, T. E. (1977) *J. Mol. Biol.* **113**, 275–293.
13. Roder, H., Elöve, G. A. & Englander, S. W. (1988) *Nature (London)* **335**, 700–704.
14. Schmid, F. X. & Baldwin, R. L. (1979) *J. Mol. Biol.* **135**, 199–215.
15. Kim, P. S. & Baldwin, R. L. (1980) *Biochemistry* **19**, 6124–6129.
16. Roder, H. & Wüthrich, K. (1986) *Proteins: Struct. Funct. Genet.* **1**, 34–42.
17. Udgaonkar, J. B. & Baldwin, R. L. (1988) *Nature (London)* **335**, 694–699.
18. Bycroft, M., Matouschek, A., Kellis, J. T. & Fersht, A. R. (1990) *Nature (London)* **346**, 488–490.
19. Karplus, M. & Weaver, D. L. (1976) *Nature (London)* **260**, 404–406.
20. Eigen, M. (1964) *Angew. Chem., Int. Ed. Engl.* **3**, 1–19.

21. Berger, A. & Linderstrom-Lang, A. (1957) *Arch. Biochem. Biophys.* **69**, 106–118.
22. Hvidt, A. & Nielsen, S. O. (1966) *Adv. Protein Chem.* **21**, 287–386.
23. Englander, S. W. & Poulsen, A. (1969) *Biopolymers* **7**, 329–339.
24. Molday, R. S., Englander, S. W. & Kallen, R. G. (1972) *Biochemistry* **11**, 150–158.
25. Bryan, W. D. (1970) *Recent Prog. Surf. Sci.* **3**, 101–120.
26. Woodward, C. K., Simon, I. & Tuchsén, E. (1982) *Mol. Cell. Biochem.* **48**, 135–160.
27. Richards, F. M. (1979) *Carlsberg Res. Commun.* **44**, 47–63.
28. Tsuboi, M. & Nakanishi, M. (1979) *Adv. Biophys.* **12**, 101–130.
29. Lumry, R. & Rosenberg, A. (1975) *Col. Int. C. N. R. S. L'Eau Syst. Biol.* **246**, 55–63.
30. Englander, S. W. & Kallenbach, N. R. (1984) *Q. Rev. Biophys.* **16**, 521–655.
31. Honig, B. H., Hubbell, W. L. & Flewelling, R. F. (1986) *Annu. Rev. Biophys. Biophys. Chem.* **15**, 163–193.
32. Wagner, G. & Wüthrich, K. (1982) *J. Mol. Biol.* **160**, 343–361.
33. Kuwajima, K. & Baldwin R. L. (1983) *J. Mol. Biol.* **169**, 299–324.
34. Wand, A. J., Roder, H. & Englander, S. W. (1986) *Biochemistry* **25**, 1107–1114.
35. Louie, G., Thao, T., Englander, J. J. & Englander, S. W. (1988) *J. Mol. Biol.* **201**, 755–764.
36. Louie, G., Englander, J. J. & Englander S. W. (1988) *J. Mol. Biol.* **201**, 765–772.
37. Wright, P. & Narula, S. S. (personal communication).
38. Wyman, J. (1967) *J. Am. Chem. Soc.* **89**, 2202–2218.
39. Englander, S. W. & Mauel, C. (1972) *J. Biol. Chem.* **247**, 2387–2394.
40. Englander, S. W. & Englander, J. J. (1983) *Structure and Dynamics: Nucleic acids and Proteins*, eds. Clementi, E. & Sarma, R. H. (Adenine Press: Guilderland, N. Y.) pp. 421–434.
41. Malin, E. L. & Englander, S. W. (1980) *J. Biol. Chem.* **255**, 10 695–10 701.
42. Englander, S. W. & Rolfe, A. (1973) *J. Biol. Chem.* **248**, 4852–4861.
43. Englander, S. W. (1975) *Ann. N. Y. Acad. Sci.* **244**, 10–27.
44. Rosa, J. J. & Richards, F. M. (1979) *J. Mol. Biol.* **133**, 399–416.
45. Englander, S. W., Calhoun, D. B., Englander, J. J., Kallenbach, N. R., Liem, R. K. H., Malin, E. L., Mandal, C. & Rogero, J. R. (1980) *Biophys. J.* **32**, 577–590.
46. Mallikarachchi, D., Burz, D. S. & Allewell, N. M. (1989) *Biochemistry* **28**, 5386–5391.
47. Englander, J. J., Rogero, J. R. & Englander, S. W. (1985) *Anal. Biochem.* **147**, 234–244.
48. Ray, J. & Englander, S. W. (1986) *Biochemistry* **25**, 3000–3007.
49. Ackers, G. K. & Smith, F. R. (1985) *Annu. Rev. Biochem.* **54**, 597–629.
50. Perutz, M. F. (1970) *Nature (London)* **228**, 726–734.
51. Baldwin, J. & Chothia, C. (1979) *J. Mol. Biol.* **129**, 175–220.
52. Pettigrew, D. W., Romeo, P. H., Tsapis, A., Thillet, J., Smith, M. L., Turner, B. W. & Ackers, G. K. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1849–1853.
53. Baldwin, J. M. (1975) *Prog. Biophys. Mol. Biol.* **29**, 225–320.
54. Imai, K. (1973) *Biochemistry* **12**, 798–807.
55. Kilmartin, J. V. & Hewitt, J. A. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 311–314.
56. Turner, G. J. (1989) PhD Dissertation, The Johns Hopkins University.
57. Imai, K., Tsuneshige, A., Harano, T. & Harano, K. (1989) *J. Biol. Chem.* **264**, 11174–11180.
58. Kilmartin, J. V., Breen, J. J., Roberts, G. C. K. & Ho, C. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1246–1249.
59. Russu, I. M., Ho, N. T. & Ho, C. (1980) *Biochemistry* **19**, 1043–1052.
60. Ohe, M. & Kajita, A. J. (1980) *Biochemistry* **19**, 4443–4450.
61. Matsukawa, S., Itatami, Y., Mawatari, K., Shimokawa, Y. & Yoneyama, Y. (1984) *J. Biol. Chem.* **259**, 11479–11486.

Polymerization of Hemoglobin S Amidated at the Contact Residue Glu 43(β)

A. Seetharama Acharya, Lakshmi Khandke¹, Brian T. Chait
and M. Janardhan Rao

Division of Hematology, Department of Medicine, Albert Einstein College of
Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA
and

¹The Rockefeller University, 1230 York Avenue, New York 10021, USA

Introduction

The polymerization of hemoglobin S in the deoxy conformation is a direct consequence of the mutation of glutamic acid residue at the sixth position of the β -chain of HbA to valine (1–3). A high degree of specificity exists in the mutation that endows this segment (A-helix) of β -chain a ‘new quinary structural memory’. Mutation of Glu 6 to glutamine, or alanine, or lysine is unable to encode the segment with this structural memory (4). Recent site directed mutagenetic studies have demonstrated that the specificity to encode the ‘new quinary structural memory’ is in the hydrophobicity of the amino acid residue at the sixth position rather than being unique to the Val residue. Mutants of HbA with Ile or Leu at this position also polymerize, their propensity to polymerize is even higher than that of HbS (5, 6). Early studies have demonstrated that the polymerization of deoxy HbS involves hydrophobic interactions (7). Subsequent studies have, however, established that electrostatic interactions also play a crucial role in the polymerization reaction (8, 9). Thus the polymerization of deoxy HbS is a complex phenomenon involving various types of noncovalent interactions of intermolecular regions that apparently generate the crucial stabilization energy to facilitate the polymerization reaction.

X-ray diffraction studies of deoxy HbS crystals have suggested that the carboxyl groups of Glu 22(β), Glu 121(β), Glu 23(α), and Asp 73(β) are present at or near one or more intermolecular contact regions. More recently Padlan and Love (10, 11) suggested that Glu 43(β) is also present at a contact region. The participation of these carboxylates in intermolecular interactions could be a consequence of a higher pK_a of these carboxylates compared to those of the others. A study of the chemical reactivity of carboxylates of HbS could provide an insight into this aspect of the structure.

Carbodiimide activated coupling of nucleophiles to the carboxyl groups of proteins

is a relatively simple and mild procedure for the derivatization of β -and/or γ -carboxyl groups of proteins (12). The derivatization of carboxylates of proteins is very efficient at moderately acidic pH, and is generally carried out at pH 4.0, a region in which most of the carboxylates of the protein are protonated. However, in view of the possible complications associated with the dissociation of Hb tetramers as well as precipitation of the protein in this low pH region of 4.0, an intermediate pH of 6.0 was selected for the derivatization of the carboxyl groups. The initial studies demonstrated that in the presence of water soluble carbodiimide, 1-ethyl-3[-3'-(dimethylamino propyl)] carbodiimide (EDC), a selective coupling of the nucleophiles to the carboxyl groups of HbS occurs around pH 6.0 (13, 14). The selectivity of the amidation reaction was very distinct with the two nucleophiles investigated. When glycine ethyl ester (GEE) was used as the nucleophile, more than 75% of the amidation was on the γ -carboxyl group of Glu 43(β). On the other hand, with glucosamine as the nucleophile, the extent of amidation was about 50% lower, as compared with that obtained using glycine ethyl ester. Besides, with glucosamine, amidation also occurred at Glu 22(β), besides at Glu 43(β). The extent of reaction at these two carboxylates were nearly the same.

Though extent of amidation with glucosamine is significantly lower than that seen with glycine ethyl ester, the solubilities of the two amidated products (unfractionated) are nearly the same (14). This was rather surprising, and casts some doubt on the implication of Glu 43(β) as the contact residue. Therefore, preparation of a homogeneous derivative of HbS amidated only at Glu 43(β) seemed essential to unequivocally establish the role of Glu 43(β) in the polymerization reaction. Accordingly the preparation of homogeneous amidated HbS has been undertaken. The isolation and characterization of Di-GEE-HbS, the oxygen affinity and solubility behaviour of this amidated derivative of HbS, has been studied now and presented here.

Materials and Methods

Modification of Carboxyl Groups of HbS

HbS was purified from erythrocyte lysate from patients with sickle cell disease using chromatography on DE-52 (15). The amidation reaction was carried out essentially by the method described earlier (13). HbS was dialyzed against 0.1 M KCl, adjusted to pH 6.0 and concentrated by ultrafiltration. [^{14}C] GEE (New England Nuclear, specific activity 52.2 mCi/mmol) was mixed with cold GEE to give the desired specific activity. The protein sample was incubated with 100 mM [^{14}C] GEE at 23 °C in a vessel attached to a Radiometer pH stat. The coupling reaction was initiated by adding appropriate amounts of 1-ethyl-3(-3'-dimethylamino propyl) carbodiimide (EDC). The reaction was maintained at pH 6.0 by automatic titration with 0.01 M HCl. After 1 hour the reaction mixture was passed through a column of Sephadex G-25 equilibrated with 10 mM phosphate buffer, pH 6.0, containing 1 mM EDTA.

Purification of Derivatized HbS

After the removal of excess reagents by gel filtration the derivatized protein was chromatographed on carboxymethyl cellulose (CM-52) equilibrated to pH 6.0 with

10 mM potassium phosphate buffer containing 1 mM EDTA. The protein was eluted using a linear gradient of 10 mM potassium phosphate buffer pH 6.0 to 50 mM phosphate buffer pH 8.3. Both buffers contained 1 mM EDTA.

Characterization of the Amidated Derivative

The procedures used for the chemical analysis of the derivative, namely, tryptic peptide mapping, amino acid analysis, and amino acid sequencing were essentially as described earlier (13, 14)

Mass Spectrometry

Mass spectrometric measurements were performed on the 252 Cf fission fragment ionization time-of-flight mass spectrometer described previously (16–18). Samples were prepared for measurement by absorption of 1 nmole of peptide from solution (0.2 nmole/ μ l in 0.1% TFA) on a thin nitrocellulose film, as previously described (18). The accuracy of the mass measurements was generally better than 200 pm.

Oxygen Equilibrium Measurements

The oxygenation curves of HbS and derivatives of HbS were recorded using an Amicon Hem-O-Scan, at pH 7.4 and 37 °C as described by Benesch *et al.* (19).

Solubility Studies

Solubility measurements were performed by the ultracentrifugation procedure as discussed by Benesch *et al.* (20). The carbon monoxide in the HbS and its amidated derivatives were removed by photolysis (21). The protein was concentrated by ultrafiltration to about 35 g/dl and this solution was used for the solubility measurements. The samples were deoxygenated, mixed well, and incubated for 2 hours after which the Hb concentration in the supernatant was measured according to the method of Van Assendelft (22).

Results

Isolation and Characterization of Di-GEE-HbS

Glu 43(β) is the major reactive site for amidation with glycine ethyl ester (13). After some preliminary analysis, reaction of HbS (0.5 mM) with 10 mM EDC in the presence of 100 mM GEE for 1 h at pH 6.0, 23 °C was chosen for derivatization and the subsequent isolation of Di-GEE-HbS (figure 1). The Di-GEE-HbS isolated was further purified by rechromatographing it on a CM-52 column (figure 1 inset). The repurified Di-GEE-HbS contained two moles of GEE per mole of tetramer.

Tryptic peptide mapping of Di-GEE-HbS showed that all the radioactivity (GEE incorporated) eluted at two positions (figure 2). One peptide eluting around 64 minutes was designated as Peak A and the other eluting around 72 minutes was designated as Peak B. Peaks A and B were isolated and rechromatographed on HPLC using an aquapore RP-300 column at pH 6.0 using an acetonitrile gradient (5 to 50%) (14).

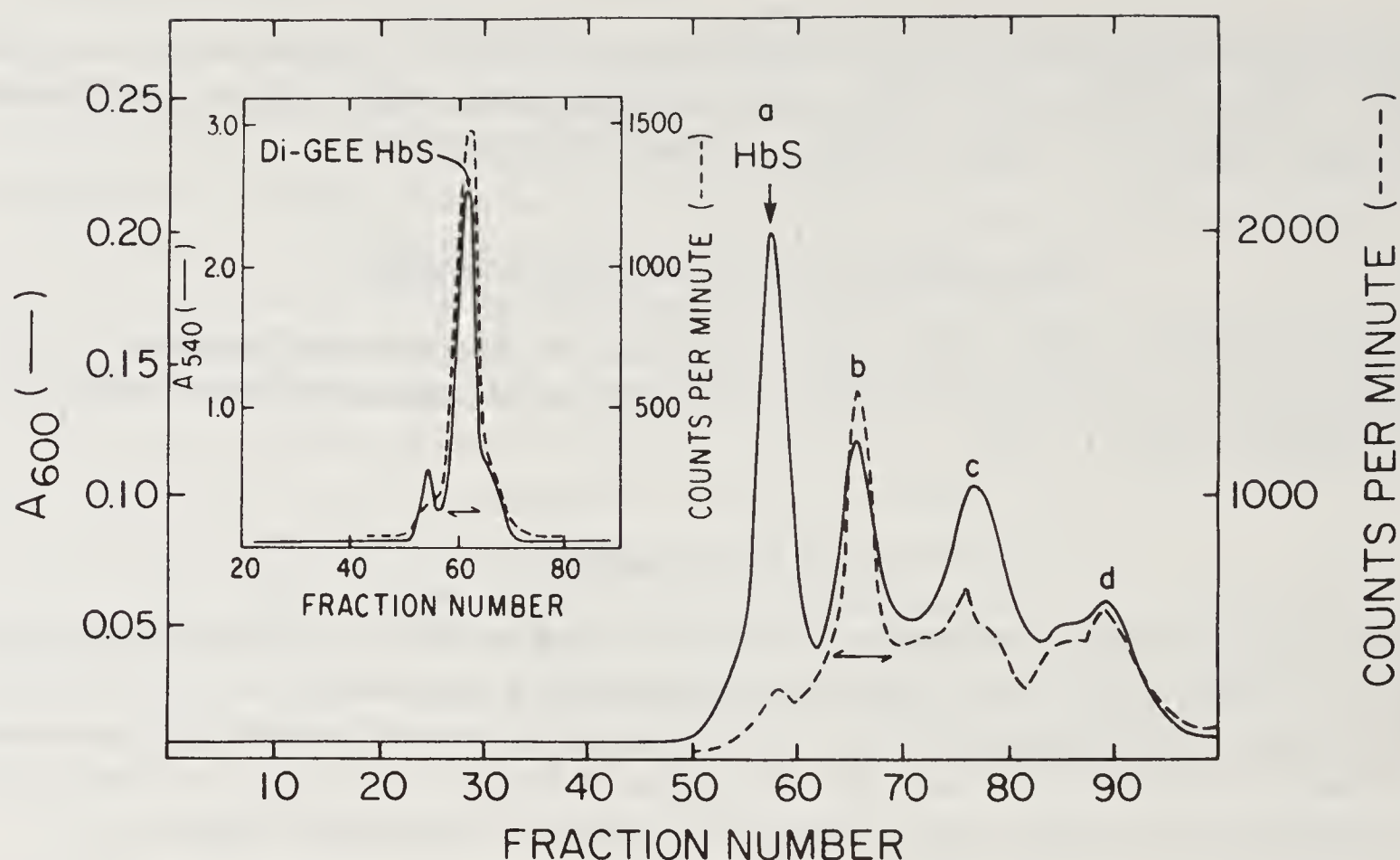


Figure 1. Preparation of Di-GEE-HbS: The excess reagents were removed from amidated HbS by passing through a Sephadex G-25 column (2.5 × 40 cm) eluted with 10 mM potassium phosphate buffer, pH 6.0. Amidated HbS was then chromatographed on CM-52 as described under materials and methods. The major radioactive fraction, eluting after HbS position and marked was pooled and concentrated. It was then dialyzed against the starting buffer, (10 mM potassium phosphate 1 mM EDTA) and rechromatographed on C-52 (0.9 × 30 mM) under the same conditions (insert). The main component marked 'Di-GEE-HbS' was pooled to isolate Di-GEE-HbS.

The amino acid composition of both the purified radioactive peptides corresponded very well with that of β T5 representing the segment 41 to 59 of β -chain, (table 1), except for the presence of an extra residue of glycine. In addition, the quantitation of the peptide using the specific activity of GEE and the amino acid composition corresponded very well, further confirming the homogeneity of the peptides. These results demonstrate that both of these peptides are formed as a consequence of the amidation of only one of the side chain carboxyl groups of β T5.

Both of these peptides were subjected to amino acid sequencing to identify the amidated carboxyl group of these peptides. The amount of ^{14}C -label released during each of the 15 cycles of the Edman degradation has been quantitated. Both peptides A and B showed the same sequence analysis and from both the radioactivity was released in the third cycle. This position corresponds to Glu 43(β). Thus, both these peptides correspond to β T5 amidated at Glu 43.

The covalent modification that results in the presence of two modified forms of β T5 in the tryptic digest of Di-GEE-HbS is not apparent from the sequence analyses. In an attempt to gain an insight into the structural differences between the two amidated β T5, the purified peptides A and B were subjected to mass spectral analysis (table 2). The predominant ion species (M^+H) of peptide B has a molecular weight of 2144.2 and is consistent with the calculated molecular weight for the β T5 amidated with glycine ethyl ester at Glu 43 as 2144.4. On the other hand, the predominant ion

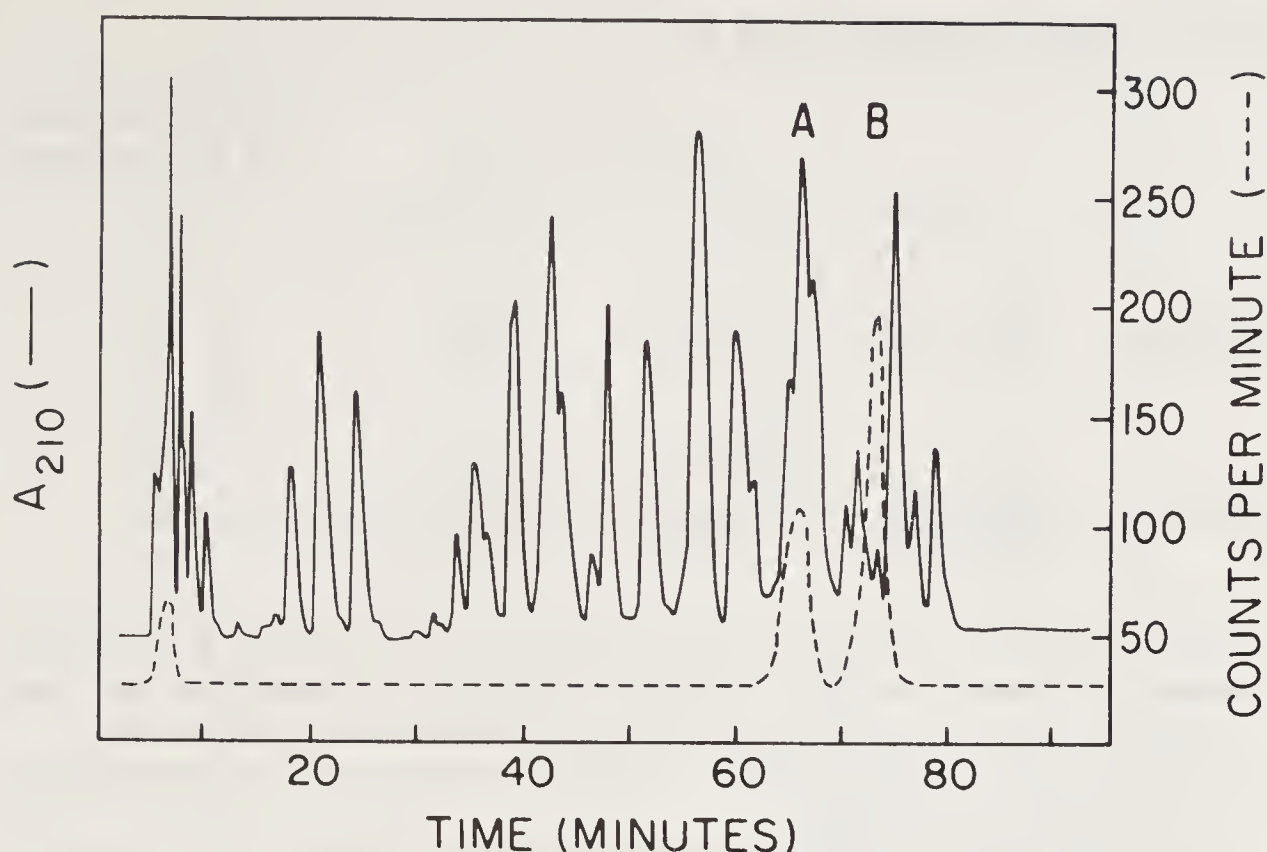


Figure 2. HPLC Analysis of the tryptic peptides of Di-GEE-HbS: The tryptic digest of Di-GEE-HbS (figure 1 insert) was taken in 500 μ l of 0.1% trifluoroacetic acid and loaded on to a Partsil 10 ODS 3 column. The peptides were eluted with a linear gradient of 5–50% acetonitrile containing 0.1% trifluoroacetic acid (Acharya *et al.* 1983) over 2 h. The column was eluted at a flow rate of 1 ml/min, and 1 ml fractions were collected. The effluent was monitored at 210 nm, 100 μ l of each fraction were used for measuring the radioactivity. The radioactive components labelled A and B were further purified by RPHPLC at pH 6.0 (reference 14) and identified as modified β T5 (corresponding to the segment 41 to 59 of β -chain of bS) by amino acid analysis and amino acid sequencing.

Table 1. Amino acid composition of amidated peptides A and B.

Amino Acid	Peptide		Expected for BT5 (segment 41 to 59)
	A	B	
Asp	3.1	3.1	3
Thr	0.8	1.0	1
Ser	1.8	2.0	2
Glu	0.9	1.0	1
Pro	1.7	2.0	2
Gly-b	2.7	2.9	2
Ala	1.1	1.0	1
Val	1.2	1.1	1
Met	0.9	0.8	1
Leu	1.1	0.9	1
Lys-a	1	1	1

a. Calculated assuming Lys as one residue.

b. β T5 contains two residues of glycine, the higher value (nearly three) of glycine in the peptides A and B represents glycine ethyl ester linked to β T5 through an isopeptide linkage.

Table 2. Mass of peptides A and B.

	Ion (a)	Molecular weight		M. W Observed
		Observed	Calculated (b)	— M· W· calculated (c)
Peptide A	(M +) +	2116·8	2144·4	— 27·6
	(M + Na) +	2116·3	2144·4	— 28·1
	(M + K) +	2117·7	2144·4	— 26·7
Peptide B	(M + H) +	2144·2	2144·4	— 0·2
	(M + Na) +	2144·1	2144·4	— 0·3
	(M + K) +	2144·4	2144·4	0·0

- a. Only the molecular weight of the major species is given. Both peptides A and B gave another component, the mass of which was higher by 16 units than the main component. This presumably corresponds to the sulfoxide form of the peptide.
- b. In calculating the molecular weight of the peptides, an average of five most abundant isotopic species is considered.
- c. The net difference in the observed molecular masses of Peptide A and Peptide B appears to be about 28 mass units.

species of the peptide A indicated that this peptide has a molecular mass of 2116·8 which is nearly 28 mass units lower than that of peptide B. A mass difference of 28 between peptides A and B could arise by the hydrolysis of the ethyl ester group of glycine ethyl ester on the γ -carboxyl of Glu 43 (loss of ethyl group-CH₂CH₃ and protonation of the carboxyl). Thus the structural studies along with the mass data of peptides suggested that peptide A was generated as a consequence of hydrolysis of the ethyl group of the glycine ethyl ester of peptide B.

The distribution of the amidated β T5 of Di-GEE-HbS as peptides A and B in the tryptic peptide map showed some dependence on the duration of the tryptic digestion. During the initial stages of the digestion, peptide B was always the major radioactive component. However with time peptide A started to accumulate. This result suggested that peptide B is accumulating at the expense of peptide A during tryptic digestion. Accordingly purified peptide B was incubated at pH 8·0 (50 mM ammonium bicarbonate) at 37°C (in the absence of trypsin) and analyzed by RPHPLC (acetonitrile-TFA system). Incubation of peptide B at pH 8·0 and 37°C for 18 h resulted in a 40% conversion of the sample (figure 3b) to peptide A. This conversion was absent in the sample incubated at 4°C. Mass spectral analysis of this new peptide formed on incubation of peptide B at pH 8·0 and 37°C confirmed a loss of a group with a molecular mass of 28 mass units. Amino acid analysis and amino acid sequencing confirmed it to be peptide A. Thus, these structural studies establish that in peptide A, the ethyl group of GEE introduced onto Glu 43(β) has been hydrolyzed, and this is a trypsin independent, temperature dependent process.

Oxygen Equilibrium Properties of the Di-GEE-HbS

The oxygen affinity of HbS and the Di-GEE-HbS was determined at pH 7·4 and 37°C (figure 4). The derivatization of HbS with GEE at Glu 43(β) results in an increase in the oxygen affinity of the molecule, the P50 decreased from a control value of about 10·0 to 7·6 on derivatization of Glu 43(β). However, the Hill coefficient was not significantly altered ($n = 2·8$) and was the same as that of the control HbS. The

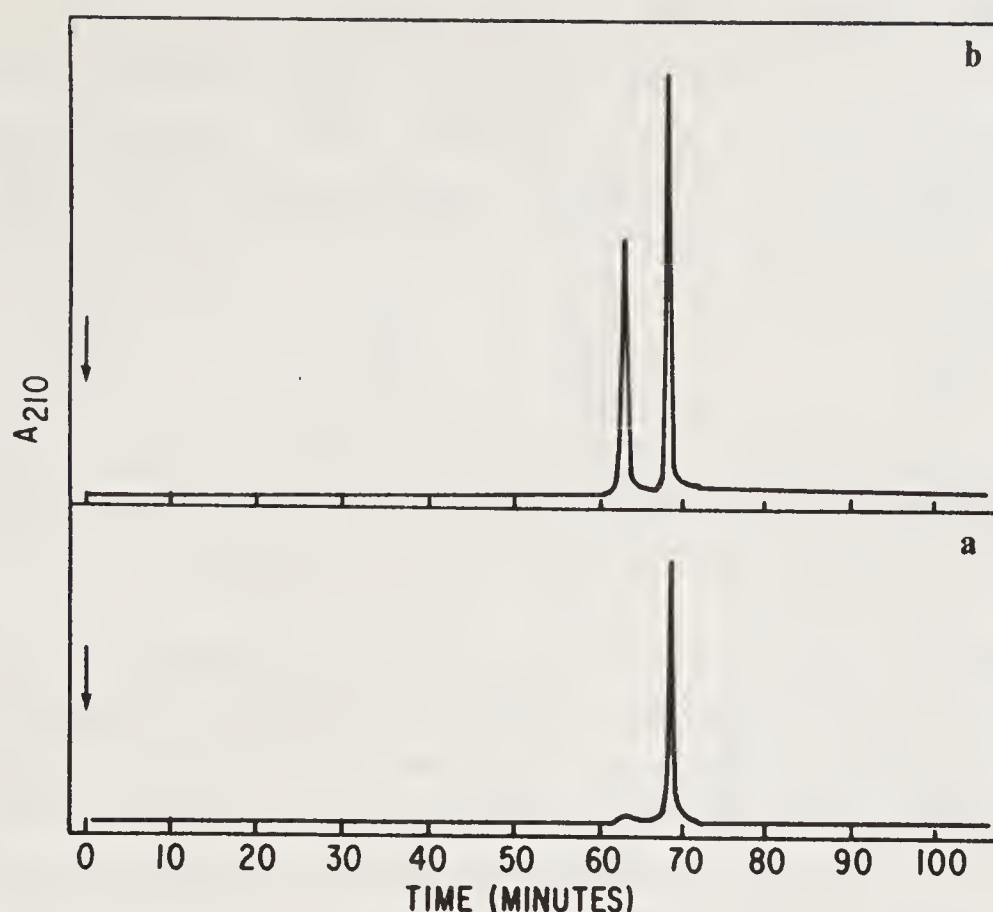


Figure 3. Hydrolysis of the ethyl ester group of amidated β T5. HPLC analysis of peptide A incubated at pH 8.0, 37 °C for 18 h. The peptide sample after incubation at pH 8.0, was lyophilized, and chromatographed on a Brownlee aquapore RP-300 column. The gradient system was the same as that used in figure 2A (control peptide A B). Peptide A incubated at pH 8.0.

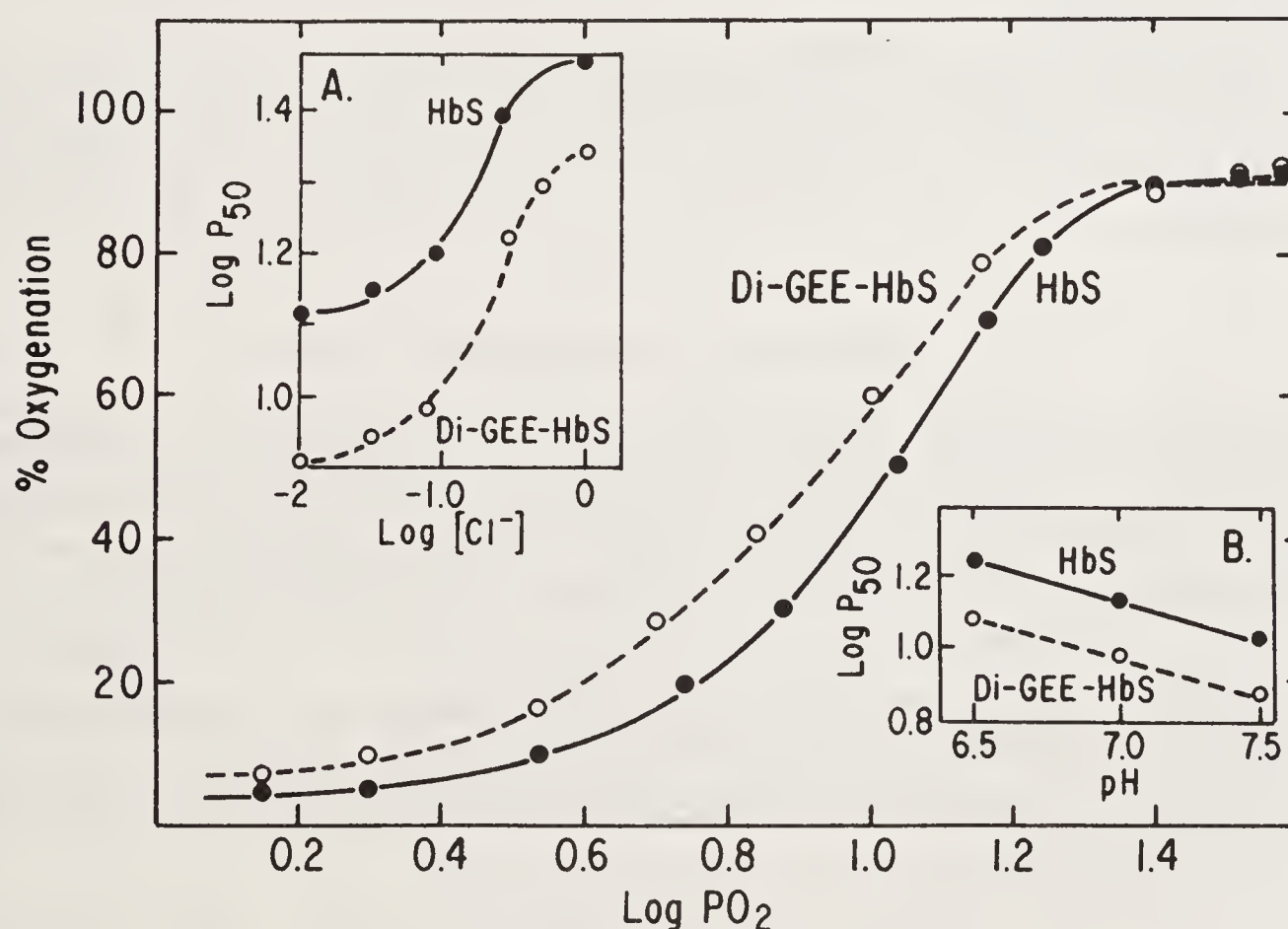


Figure 4. Oxygen affinity of Di-GEE-HbS: Oxygen equilibrium curves of HbS and amidated HbS at pH 7.4 (50 mM bis-Tris) and 37 °C. The curves were recorded in an Amicon Hem-O-Scan. A) HbS, B) HbS amidated at Glu 43(β). Insert A shows the influence of chloride ion on the oxygen equilibrium properties of Di-GEE-HbS. The samples prepared as described in the text were mixed with varying amounts of chloride and the oxygen equilibrium curves were determined in the Hem-O-Scan. The inset B presents the O₂ equilibrium properties of Di-GEE-HbS and HbS as a function of pH.

Table 3. Oxygen affinity of amidated HbS.

Modulators of O ₂ affinity	O ₂ affinity P50		P50 of Di-GEE P50 HbS
	HbS	Di-GEE-HbS	
None	10.0	7.6	0.76
DPG	26.5	18.5	0.70
Chloride	30.5	22.5	0.73

The oxygen affinity of hemoglobin S and Di-GEE-HbS was measured using hemoscan in 50 mM bis-Tris., pH 7.4. There was 2 fold molar excess of DPG to hemoglobin and 2000 fold excess of chloride to hemoglobin.

derivatization of Glu 43(β) does not influence the cooperativity in binding of oxygen to the protein. Besides, the binding of DPG to Di-GEE-HbS leads to a decrease in the oxygen affinity of the protein to approximately the same degree (table 3) as that seen with unmodified HbS. The Hill coefficient of the amidated HbS in the presence of DPG was also 2.8. The results suggest normal subunit interaction in this derivatized HbS in the presence of DPG as well.

The oxygen affinity of Di-GEE-HbS has also been determined as a function of chloride concentration and compared with that of unmodified HbS (figure 4, inset A). The overall shape of curves (a plot of P50 versus chloride) for HbS and Di-GEE-HbS is nearly the same. This demonstrates that modulation of quaternary interaction of HbS by the chloride ion is also not significantly influenced by the amidation of Glu 43(β).

The oxygen affinity of Di-GEE-HbS has also been studied as a function of pH in order to determine whether amidation of Glu 43(β) influences the alkaline Bohr effect of HbS. The value of log P50/pH was 0.52 for HbS, and 0.50 for the Di-GEE-HbS (figure 4, inset B).

Solubility of Di-GEE-HbS

The equilibrium solubilities of HbS, and purified Di-GEE-HbS were determined by the procedure of Benesch *et al.* (20). The solubility of the Di-GEE-HbS was nearly 25% higher than that of the control HbS (the solubility increased from a control value of 16 g/dl to 20.5 g/dl). Though the solubility of the Di-GEE-HbS was increased by nearly 25%, the concentration of the protein in the gel was still 45 g/dl, which is the same as that for the control HbS sample. This suggests a normal fiber structure is maintained even after the Glu 43(β) of HbS is derivatized. The increased solubility of Di-GEE-HbS, thus suggests the participation of the γ -carboxylate of Glu 43(β) in the polymerization process.

Discussion

The isolation of Di-GEE-HbS is consistent with the high chemical reactivity of the γ -carboxylate of Glu 43(β) (13,14 and 23). However, the β T₅, from the Di-GEE-HbS elutes as two chromatographically distinct components from RPHPLC columns, in much the same way as we had seen earlier in the tryptic map of

unfractionated amidated HbS (13,23). This casts some doubt on the molecular homogeneity of the Di-GEE-HbS that has been isolated. This prompted us to undertake the detailed mass spectral analysis of the two samples of amidated βT_5 from Di-GEE-HbS. The results have confirmed that the hydrolysis of amidated βT_5 during the tryptic digestion is responsible for the anomalous behavior of amidated βT_5 on the RPHPLC columns. These results are also consistent with the earlier observation that when the amidation is carried out with glucosamine (14), the amidated βT_5 elutes as a single molecular component.

All the hydrolytic reaction detected by RPHPLC appears to have occurred only during the tryptic digestion of Di-GEE-HbS. When Di-GEE-HbS is maintained at pH 7.5, and 37°C for periods up to 3 h, no detectable hydrolysis of the ethyl groups of the amidated protein occurred (CM cellulose chromatographic behavior of Di-GEE-HbS remains unchanged). It may be added here that though hydrolysis of ethyl and methyl ester groups on the side chain carboxyl groups of RNase-A is known to occur around pH 10.5 (24, 25), very little of such a hydrolysis occurs around pH 8.0. This can be considered as a reflection of the fact that, the observed hydrolysis is facilitated once the amidated sample is converted into the tryptic peptides. The observation that the pure amidated βT_5 containing the ethyl ester group, is converted to one without the ethyl ester group, in an overall yield of only 40% even after an incubation for a period of 18 h at 37°C, demonstrates that this hydrolysis occurs at a very slow rate.

It is not clear whether this hydrolysis is a consequence of some unique structural aspects of the amino acid sequence of βT_5 . Though the hydrolysis of the ethyl ester group by itself is not a surprising result, the fact that such a phenomenon has not been observed previously in protein structural studies is surprising. The nucleophiles, glycine ethyl ester and glycine methyl ester have been extensively used in the amidation studies of proteins since the introduction of this procedure for the chemical modification of the carboxyl groups of proteins by Hoare and Koshland (12). The higher resolution of the HPLC itself does not appear to be a factor for detection of this hydrolytic reaction in our studies. Huynh (26), has amidated 5-enolpyruvyl shikimate 3 phosphate synthase with EDC and GEE. However, a hydrolysis of the ethyl ester groups of the amidated protein/peptide has not been detected. Thus the present demonstration of the hydrolysis of the ethyl ester group of GEE from the amidated peptide, should be considered as a cautionary note in the chemical modification studies of carboxylates of proteins using glycine ethyl ester as a nucleophile.

The cooperative binding of oxygen to Di-GEE-HbS and the modulation of oxygen affinity by diphosphoglycerate (DPG), as well as by chloride, clearly suggests that its quaternary structure is nearly the same as that of HbS. However, the solubility of HbS is increased nearly 25% on amidation implicating a structural role for the γ -carboxyl of Glu 43(β) in the polymerization reaction. The concentration of the protein in the gel of Di-GEE-HbS is nearly the same as that of control HbS sample (45 g/dl). These results reflect the overall similarity in the polymer structure of Di-GEE-HbS to that formed from the unmodified HbS.

Glu 43(β) is a part of the CD region of β -chain that makes a portion of the heme pocket. However Glu 43(β) itself contributes little to the organization of this hydrophobic pocket that stabilizes heme in its position. The hydrophilic side chain of Glu 43(β) extends towards the surface of the hemoglobin molecule. Thus the

modification of γ -carboxylate of Glu 43(β) is unlikely to affect the heme pocket and the stability of the molecule. The limited influence of amidation on the overall quaternary structure of HbS is consistent with these structural aspects of the protein. Two mutations of Glu 43(β) have been reported. In Hb Hoshida, Glu 43(β) is mutated to Gln, (27) and in Hb G Galveston/Hb G Port Arthur/Hb G Texas Glu 43(β) is mutated to Ala (28). Both are stable and have oxygen affinities comparable to that of Hb. Thus Di-GEE-HbS is comparable to these mutant forms of Hb in terms of its structure and stability.

Fermi and Perutz (29) have indicated that Glu 43(β) could form a salt bridge with Arg 92(α). Thus mutation of Arg 92(α), could perturb the microenvironment of Glu 43(β). Two mutant forms of Hb, with a mutation of Arg 92(α) are known, namely Hb Chesapeake and Hb Capetown. In Chesapeake Arg 92(α) is mutated to a Leu residue, whereas in Hb Capetown it is replaced by a glutamine residue. The perturbation of the Glu 43(β) microenvironment in Hb Chesapeake results in an increased oxygen affinity for the protein and the loss of cooperativity ($n = 1.3$). Hb Capetown also shows an increased oxygen affinity. However, the increase in the oxygen affinity in the case of Hb Capetown is not as large as that of Hb Chesapeake. Besides the cooperativity of Hb Capetown is normal, unlike that of Hb Chesapeake. A 2-fold increase in oxygen affinity of Hb Capetown compared to that of HbA was observed, whereas that of Hb Chesapeake is increased by about eightfold (30). The cooperativity of Di-GEE-HbS is normal, and the oxygen affinity of Di-GEE-HbS is increased only slightly suggesting that the perturbation at the microenvironment may be comparable to that occurring in Hb Capetown.

The intermolecular contact regions of deoxy HbS that stabilizes the HbS fiber have been broadly grouped into three classes (a): inter double strand (b) intra double strand axial and (c) intra double strand lateral (10, 11, 31 and 32). Edelstein (33) was the first one to propose the Glu 43(β) as an inter double strand contact residue based on his model building studies. Subsequently Padlan and Love (10, 11) have investigated the possible molecular interactions in the deoxy HbS fiber, and have suggested the molecular interactions of Glu 43(β) with Ala 53(α) and Glu 54(α). Watowich *et al.* (31) have synthesized a model for the structure of clinically relevant HbS fiber by combining the X-ray crystallographic coordinates of HbS molecules (10, 11) with the three-dimensional reconstruction of electron micrographs of HbS fibers. In their model, the inter double strand contact region of β chain appears to be concentrated along the helices D and E; and turns CD and FG. However the mutations of this region of β chain for which the influence on polymerization have been investigated, has no influence on the polymerization, i.e., an apparent discrepancy exists between the model and the solution studies (32). However, by and large, only the influence of the mutation on the *trans* position could be investigated. The exception is the residue Pro 58 for which the influence in the *cis* position has also been investigated. However, it also has no influence. Amidation of the Glu 43(β), is the first perturbation of this region that has shown a solubilizing influence.

An aspect of considerable interest, is the high reactivity of Glu 43(β) for amidation. This high reactivity is not specific for HbS, HbA also reacts in nearly the same fashion (Khandke and Acharya, unpublished results). The reactivity of Glu 43(β) is not unique for the tetrameric structure. Glu 43(β) exhibits a high reactivity even in the isolated β^s chain (23). The higher reactivity of the γ -carboxyl group Glu 43(β) for amidation at pH 6.0 reflects a higher propensity of this carboxyl to be in the protonated

form. The pK_a of this γ -carboxyl has been estimated to be 6.35 (Rao and Acharya, manuscript in preparation). Though the significance of higher pK_a of γ -carboxyl of Glu 43(β) or the structural basis for this higher reactivity is not clear at this stage, a study of the reactivity (or changes of it) of this carboxyl in the deoxy conformation and during the early phases of polymerization should be very informative. The high reactivity of Glu 43(β) should certainly prove useful to introduce other 'structural probes' into this contact region for monitoring the conformational aspects during polymerization. Photo affinity label could also be introduced at this contact site, and this could be used to map the complementary site of this contact region by the cross linking approach.

One of the inter double strand intermolecular contact residues that has been investigated in great detail is Asp 47(α). Mutation of this residue to His as in Hb Sealey, increased the solubility of HbS. Besides the gels from such a molecular variant of HbS is 'soft' and has an altered fiber structure (34, 35). Asp 47(α) is in the CD region of α -chain, in much the same way as Glu 43(β) is a residue of CD region of β chain. However the concentration of Hb in the polymer of Di-GEE-HbS is nearly the same as the control, and this suggests a normal fiber structure. In view of the fact that Glu 43(β) and Asp 47(α) represent the inter double strand contact residues, it should be of interest to generate molecular variants of HbS with these two perturbations, and establish the synergy and/or additivity (36) of the solubilizing influence of two inter double strand contact regions.

Acknowledgements

This research is supported by NIH Grant HL-27183, HL38655, and Grant-in-Aid from the American Heart Association, New York City Affiliate. ASA was an Established Fellow of the American Heart Association, New York City Affiliate. The authors wish to thank Dr. Lennette J. Benjamin for the supply of blood from patients with sickle cell anemia. The technical assistance of Youngnan J. Cho and Leslie Sussman are greatly acknowledged. We thank Dr. R. E. Benesch for her encouragement in the equilibrium measurement of the solubility and S. Kwong for her help in these measurements. Protein sequence analysis was performed by the Rockefeller University Protein Sequencing Facility (supported in part by funds provided by the U. S. Research Office for the purchase of the equipment). The mass spectrometric measurements were made by the Rockefeller University Mass Spectrometer-Biotechnology Research Resource supported by the Division of Research Resources, NIH (RR-00862).

References

1. Dean J. & Schechter A. N. (1978) *N. Engl. J. Med.* **299**, 752–763.
2. Dean J. & Schechter A. N. (1978) *N. Engl. J. Med.* **299**, 807–811.
3. Dean, J. & Schechter, A. N. (1978) *N. Engl. J. Med.* **299**, 863–870.
4. Nagel, R. L., Johnson, J., Bookchin, R. M., Garel, M. C., Rosa, J., Schiliro, G., Wajeman, M., Labie, D., Moo-Penn, W. & Castro, O. (1980) *Nature (London)* **283**, 832–834.
5. Baudin-Cheih, V., Pagnier, J., Marden, M., Bohn, B., Kister, J., Schaad, O., Edelstein, S. J. & Poyart, C. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1845–1849.

6. Adachi, K., Rapport, E., Eck, H. S., Konitzer, P., Sims, J. Y. & Sussey, S. (1990) Abstracts, 15th Annual Meeting of National Sickle Cell Center, pp. 2.
7. Murayama, M. (1966) *Science* **153**, 145–149.
8. Bookchin, R. M. & Nagel, R. L. (1973) in *Sickle Cell Disease*, ed. Abramson, H., Bertles, J. F., & Weathers, D. L. (C. V. Mosby. Co., St. Louis).
9. Benesch, R. E., Kwong, S., Benesch, R. & Edalji, R. (1977) *Nature (London)* **296**, 772–775.
10. Padlan, E. A. & Love, W. E. (1985) *J. Biol. Chem.* **260**, 8272–8279.
11. Padlan, E. A. & Love, W. E. (1985) *J. Biol. Chem.* **260**, 8280–8291.
12. Hoare, D. G. & Koshland, D. E. (1967) *J. Biol. Chem.* **242**, 2447–2453.
13. Seetharam, R., Manning, J. M. & Acharya, A. S. (1983) *J. Biol. Chem.* **258**, 14810–14815.
14. Acharya, A. S. & Seetharam, R. (1985) *Biochemistry* **24**, 4885–4890.
15. Acharya, A. S. & Manning, J. M. (1980) *J. Biol. Chem.* **255**, 1406–1412.
16. Chait, B. T., Agosta, W. C. & Field, F. H. (1981) *Int. J. Mass Spectrom. Ion Phys.* **39**, 339–366.
17. Chait, B. T., Gisin, B. F. & Field, F. H. (1982) *J. Am. Chem. Soc.* **104**, 5157–5162.
18. Chait, B. T. & Field, F. H. (1986) *Biochem. Biophys. Res. Commun.* **134**, 420–426.
19. Benesch, R. E., Edalji, R., Kwong, S. & Benesch, R. (1978) *Anal. Biochem.* **89**, 162–173.
20. Benesch, R. E., Kwong, S., Edalji, R. & Benesch, R. (1979) *J. Biol. Chem.* **254**, 8169–8172.
21. Nigen, A. M., Nijikan, N., Lee, C. K. & Manning, J. M. (1974) *J. Biol. Chem.* **249**, 6611–6616.
22. Van Assendelft, O. W. (1970) *Spectrophotometry of Hemoglobin Derivatives* (Royal Vangorium Assen, The Netherlands).
23. Acharya, A. S. & Khandke, L. (1989) *J. Prot. Chem.* **8**, 231–237.
24. Acharya, A. S., Manjula, B. N. & Vithayathil, P. J. (1975) *Biochem. J.* **173**, 821–830.
25. Broomfield, C. A., Riehm, J. P. & Scheraga, H. A. (1965) *Biochemistry* **4**, 751–759.
26. Huynh, G. K. (1989) *J. Biol. Chem.* **263**, 11631–11635.
27. Iuchi, I., Veda, S., Hidaka, K. & Shibata, S. (1978) *Hemoglobin* **2**, 235–247.
28. Bowman, B. H., Oliver, C. P., Barnett, D. R. & Cunningham, J. R. (1984) *Blood* **23**, 193–199.
29. Fermi, G. & Perutz, M. F. (1981) in *Hemoglobin and Myoglobin Atlas of Molecular Structures in Biology*, eds. Phillips, D. C. & Richards, F. M. (Clarendon Press, Oxford, England), p. 101.
30. Nagel, R. L., Gibson, Q. H. & Jenkins, T. (1971) *J. Mol. Biol.* **58**, 643–650.
31. Watowich, S. J., Gross, L. J. & Josephs, R. (1989) *J. Mol. Biol.* **209**, 821–828.
32. Wishner, B. C., Ward, K. B., Lattman, E. E. & Love, W. E. (1975) *J. Mol. Biol.* **98**, 179–194.
33. Edelstein, S. J. (1981) *J. Mol. Biol.* **150**, 557–575.
34. Benesch, R. E., Kwong, S. & Benesch, R. (1982) *Nature (London)* **299**, 231.
35. Crepeau, R. H., Edelstein, S. J., Szalay, M., Benesch, R. E., Benesch, R., Kwong, S. & Edalji, R. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1406–1410.
36. Acharya, A. S., Sussman, L. G. & Seetharam, R. (1985) *J. Prot. Chem.* **4**, 215–225.

Crystallin Structure and Eye Lens Function

Manni Luthra, C. Sivakama Sundari, P. Guptasarma
and D. Balasubramanian*

Centre for Cellular and Molecular Biology, Uppal Road,
Hyderabad 500 007, India

The eye lens is a tissue that is comprised for the most part of fibre cells, with an outer layer of epithelial cells. These latter cells metabolize and divide in the usual fashion, and differentiate to produce long fibre-shaped cells which constitute most of the lens. The differentiated fibre cells are devoid of nuclei and other organelles and hence metabolize and turn over extremely slowly. In effect, the lens fibre cells are long-lived sacs containing cytoplasms rich in protein molecules called the crystallins, which are as old as the lens or the individual organism itself. The lens is thus a tissue which is made up largely of the crystallins (35% by weight, much of the remainder being water), which has to function throughout the lifetime of the individual, as the device that along with the cornea, focusses incident light on to the retina (1, 2).

In order to do so, the lens has to meet several requirements:

- (i) It needs to be scatter-free, which means that it cannot have particulate matter of size greater than one-tenth of the wavelength of incident light, i.e. about 40 nm;
- (ii) It needs to remain transparent to light of all wavelengths in the visible region (400–750 nm) in an equitable and efficient manner, so that there is no inherent colour bias in the lens, and
- (iii) It cannot be a rigid solid body; the lens needs to be a pliable body, easily deformed under mild pressure, with the capability of regaining its original shape upon release of the pressure. This pliability and elasticity is required in order to change the focal length of the lens so that it focusses near and far objects equally well on the retina. This property of the lens is referred to as accommodation. All these requirements are met admirably well by the eye lens which is a highly concentrated gel of the crystallins in water, with short range order rather than a rigid periodic lattice, and with no components that impede or filter light transmission in the visible or the ultraviolet A region (UV-A) of 320–400 nm(3) Any compromise in these qualities of the lens can lead to cataractogenesis, colour bias in perception and presbyopia, respectively.

The power of accommodation of the lens also seems to be influenced by the presence of a radial gradient of refractive index through the body of the lens, which may relate to the presence of a similar gradient observed with regard to protein concentrations.

* For correspondence

Discontinuities and changes in the refractivity gradient which could adversely affect light scattering are seen to accompany changes in lens transparency.

The state of transparency of the lens would thus seem to depend on the maintenance of the high density packing and solubility of the crystallins in the lens cell matrix. Both of these can be influenced by changes in the organization, structure and charge profiles of the proteins, particularly in a situation where damage resulting in such changes tends to accumulate rather than be removed by turnover. We have been interested, for the last several years, in processes which could cause the above changes, and give here a perspective account of some of our studies on the molecular changes that occur to the crystallins during cataractogenesis.

Chromophores and Fluorophores in the Lens

The three aromatic amino acid residues present in the crystallins, i.e., phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp) absorb and attenuate light in the UV-B region (280–320 nm), while these and the peptide groups themselves absorb in the UV-C region below 280 nm. Figure 1 shows that the cornea effectively cuts off all light below 300 nm. Of the light that is incident above this wavelength, the lens itself filters much of the light below 320 nm from the retina (4). This filtering is essentially because of the high crystallin concentrations which, we estimate, would lead to an optical density of the intact lens which would be far greater than 10^2 in the 280 nm region and greater than 10 at the wing of the band around 310 nm.

Since the crystallins do not have any chromophores (or fluorophores) other than the Trp, Tyr, Phe and the backbone peptide groups, the lens should be colourless. This is actually so in the rat, rabbit (and bovine) lenses, as the transmission and fluorescence spectra in figure 2 reveal. The normal, healthy human lens, however,

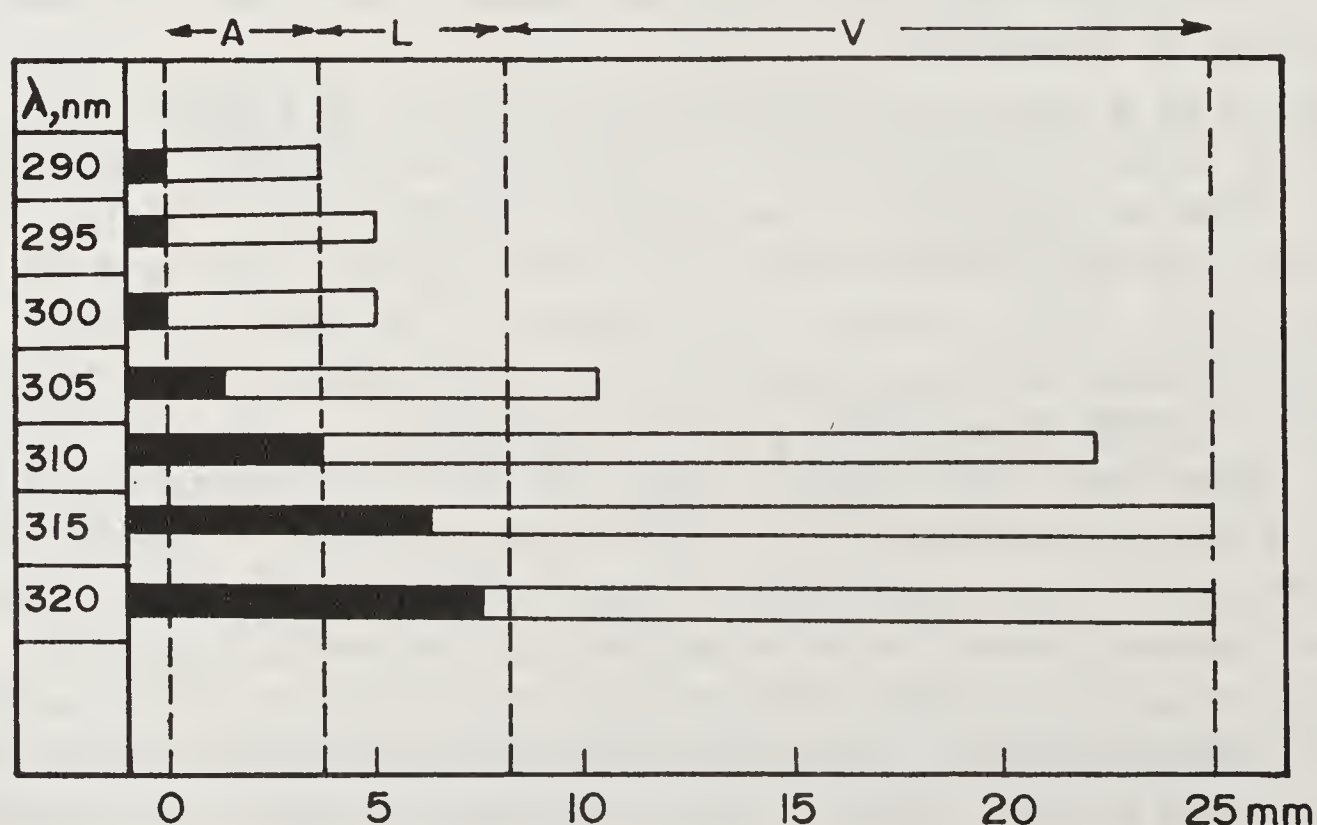


Figure 1. Penetration of UV-B light into the eye. The distances are from the cornea inwards towards the retina. A refers to the aqueous humor, L to the lens, and V to the vitreous humor. The darkened bars indicate the 10% irradiance depths while white bars correspond to 1% irradiance depths (adapted from ref. 4).

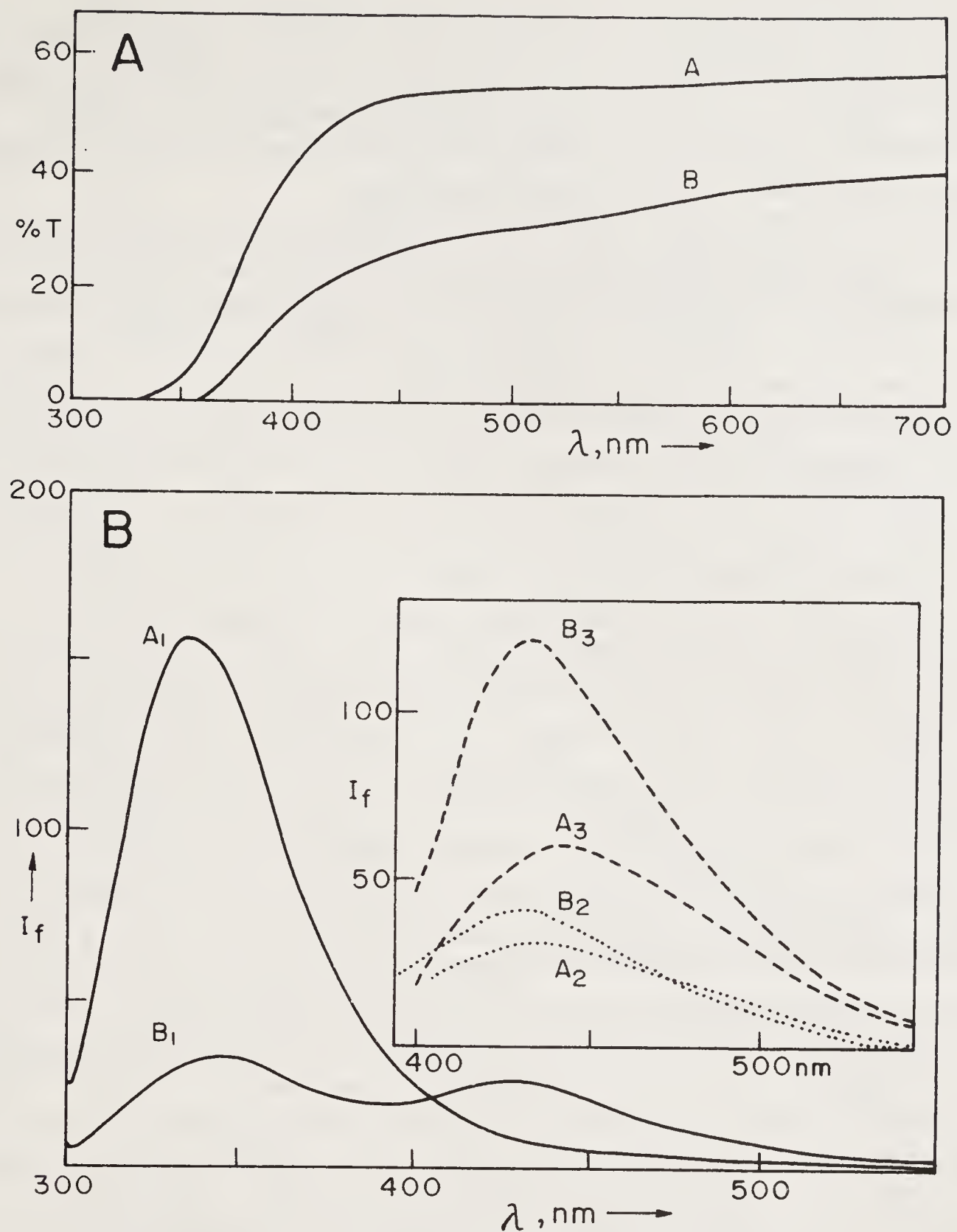


Figure 2. *Panel A.* The *in situ* transmission spectrum of a freshly excised rabbit eye lens, before (A) and after 120 minutes of irradiation with 300 nm light (B). The lens was mounted on a specially fabricated sample holder and placed in the cuvette holder of a Hitachi 330 spectrophotometer and the transmission spectrum run against air as reference. Irradiation with 300 nm light was done using the light source of a Hitachi F-4000 spectrofluorimeter, by placing the lens in the solid state sample holder accessory, placed in the sample compartment of this instrument, % T refers to the percentage transmission of light.

Panel B. *In situ* fluorescence spectra of the freshly excised rabbit lens mentioned above, before (A) and after irradiation with light of wavelength 300 nm, and bandpass of 20 nm (B). A_1 and B_1 are emission spectra obtained from excitation at 295 nm, A_2 and B_2 from excitation at 320 nm, and A_3 and B_3 from excitation at 360 nm. I_f refers to the intensity of fluorescence, in arbitrary units.

has a light straw yellow hue (as do the simian and the squirrel lenses), with an absorption band around 360 nm and fluorescent emission in the 430 nm region. This has been identified as due to 3-hydroxykynurenine glucoside (3HKG) which is thought to act as a light filter for the retina (5). We will revert to this point a little later.

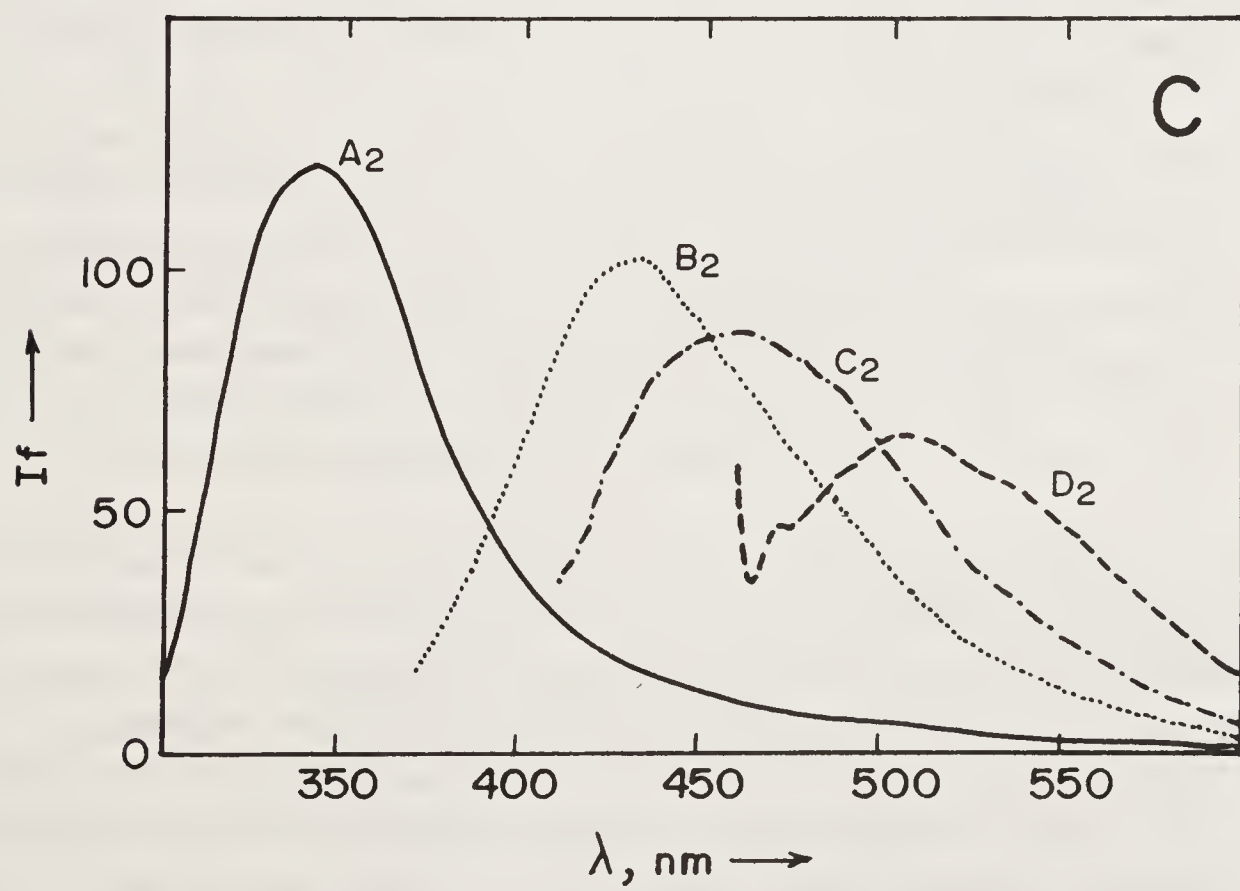
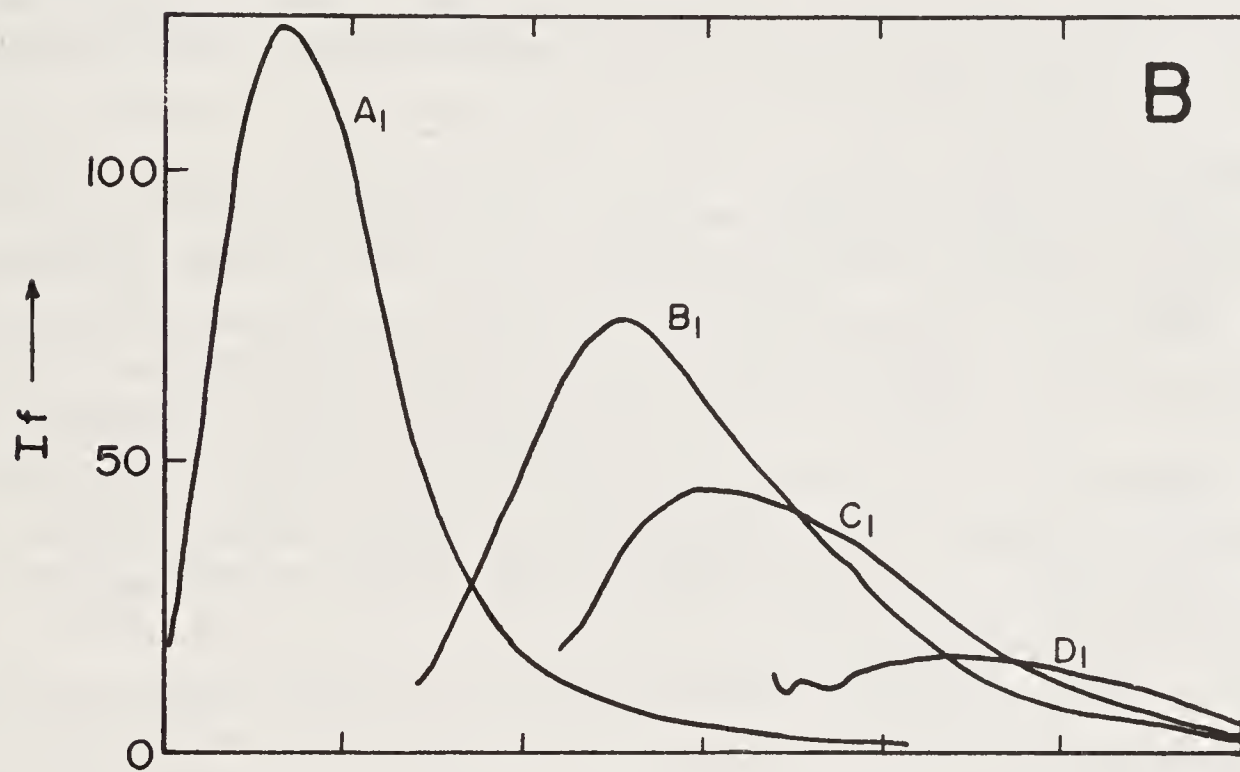
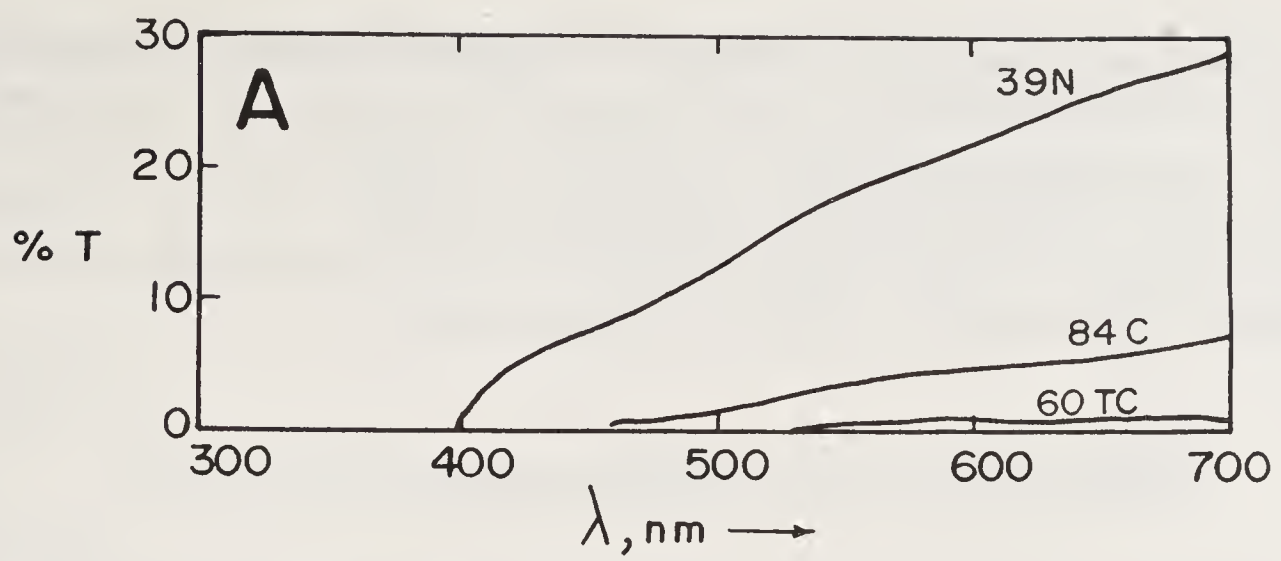
Figure 2 also shows two other important things:

- (i) When the freshly excised rabbit lens is irradiated, its emission near 335 nm is reduced and a new band around 435 nm is seen (excitation maximum 320 nm).
- (ii) In addition, the transmission characteristics of the lens become poorer due to increased light scattering. We had monitored such photolytic changes occurring in the lens *in situ*, and showed them to occur due to the photo-oxidation of the Trp residues of the crystallins (6). The resultant products of the kynurenine (Ky) family such as N-formylkynurenine (NFK), 3-hydroxykynurenine (3HK) and its glucoside 3HKG, and kynurenic acid (KUA) are thought to accumulate in the ageing human lens, occurring in the protein bound form as modified side chain groups of the proteins in the lens, as well as in the free form (7, 8).

These compounds extend the absorption range of the lens into the UV-A region, since they absorb in the 320–400 nm region. They are also apprehended to act as sensitizers which, upon absorption of light in this region, produce radicals and active oxygen species such as singlet oxygen ($^1\text{O}_2$), superoxide and hydroxide radicals (O_2^- , $\cdot\text{OH}$) and peroxide (H_2O_2), which can react with the lipids and proteins of the lens and cause covalent modifications (9). With the crystallins, these occur as protein crosslinks and high molecular weight products. Aged normal human lenses and cataractous lenses are known to contain higher proportions of high molecular weight crystallins, and modified chromophores (1, 2). Figure 3 shows the age-dependent and cataract-associated loss in the transparency of the human lens, and also the presence of new chromophores there. Some of these chromophores are in fact the various kynurenines discussed above. In addition, the presence of anthranilic acid (AA) and derivatives is also likely, particularly because AA has been detected in cataractous lenses (10, 11), and also AA appears more abundant in cataract lenses from Pakistan than in those from Europe (11). Relevant to this is the observation that tropical regions tend to have greater incidence of brunescant and nuclear cataracts than those from the temperate zones (12, 13) which might well reflect photochemical processes in the lens that could be more prevalent in the tropical sunshine belt, where cloud cover is on an average lower (14). Taylor has recently reviewed some biological effects of UV-B in particular on the eye (15).

Figure 3. *Panel A.* Transmission spectra of human lenses obtained in the manner mentioned in figure 2. The numbers 39, 84 and 60 refer to the ages of the lenses, and the characters N, C, and TC denote normal, cataractous and totally cataractous states of the lenses.

Panel B and C. Fluorescence emission spectra of normal (39 yr old) and cataractous (84 yr old) human eye lenses respectively, obtained in the manner mentioned in the legend to figure 2. A_1, B_1, C_1 and D_1 refer to emission from wavelengths 295, 353, 394, and 460 nm respectively which correspond to the excitation maxima of the fluorophores shown. A_2, B_2, C_2 and D_2 refer to emission from wavelengths 295, 351, 395 and 453 nm, corresponding to the excitation maxima of the fluorophores shown. Human lenses were obtained by the courtesy of the L. V. Prasad Eye Institute of Hyderabad.



Thus, the questions of interest in studying the molecular aspects of cataractogenesis are: (i) how are the coloured compounds produced in the lens and what role do they have in the time dependent changes that occur in the lens function, and (ii) how do protein crosslinks and high molecular weight products form and cause increased light scattering in the lens? We have been concerned with these questions in our laboratory and highlight some results that we have obtained below.

Generation of New Chromophores in the Lens

Since the lens has little or no turnover, new chromophores are generated by the modification of existing ones—largely Trp, Tyr and Phe. Much of the modification is oxidative and/or radical mediated, and produces products that invariably absorb and emit at longer wavelengths, largely in the UV-A and visible region. Modifications are mediated in the presence of light or in the dark. In the photochemical mode, the reaction can occur through direct photolysis when the Trp or Tyr residues absorb light in the 260–300 nm region and reach the excited state from where they react. Curves B₁, B₂ and B₃ of figure 2 show the generation of the kynurenines (largely NFK) upon the direct photolysis of the Trp residues of the constituent proteins (300 nm was chosen so as to selectively irradiate Trp and not Tyr or Phe). It is likely that irradiation at 320 nm (the new absorption band due to NFK) will induce further photolysis of the NFK to produce Ky, KUA and AA, but we did not see it in this instance despite such attempts (6).

The photo-oxidation of Trp to the kynurenines has been studied in some detail by the Japanese group (16), Wasserman (17) and others. The indole ring of Trp is opened and oxygen insertion occurs under aerobic condition, usually involving active species of oxygen or oxyradicals. These oxyradicals are themselves produced due to the excited state reactions of Trp. The reaction products NFK, Ky, KUA, 3HK and perhaps AA can also act as photo-sensitizers and generate oxyradicals which, in turn, cause protein modifications. We have evaluated the photodynamic ability of these compounds in our laboratory and we discuss this aspect in a later section.

Such oxidation can occur just as well in the “dark” too, provided oxyradicals are available. We have studied the reaction of Trp residues in proteins with independently generated ¹O₂ and ·OH and find the oxidation to proceed equally well (18, 19). McNamara and Augusteyn (20) have seen the oxidation of Trp to NFK in human lenses.

The other residue of interest is Tyr, which can be modified to produce dihydroxyphenylalanine (Dopa) and further into melanin type of brown pigments (21), or to produce covalently crosslinked bityrosine (Bityr) (22) which absorbs in the 325 nm region and fluoresces blue (415 nm). There is a report in literature about the occurrence of Bityr in cataractous lenses (23). We have monitored, in solution, the oxidation of Tyr residues in proteins in general and the crystallins in particular, both by direct excitation of the residue and by other means using isolated oxyradicals (6, 18, 19). We find Tyr is oxidized, as seen by a loss in its fluorescence, but we do not see the formation of Dopa or Bityr in any of these experiments (see table 1). Bityrosine is known to form in biological systems when Tyr is enzymatically oxidized by H₂O₂ in the presence of horseradish peroxidase (22), also, one expects Bityr formation to occur between two Tyr residues that are disposed in appropriate

Table 1. New chromophores obtained upon oxidation of residues in crystallins*

Chromophores	By direct photo- lysis	Through Photo- dynamics	By oxidation using pure			Present in cata- ract lens
			$^1\text{O}_2$	$\cdot\text{OH}$	H_2O_2	
Kynurenines	+	+	+	+	+	+
Anthranilic acid	—	—	—	—	+(10)	+(10, 11)
Bityrosine	—	—	—	—	? (22)	? (23)
Dopa	—	+(21)	—	—	?	—
Dopachrome						

*Numbers in brackets refer to references cited in the text

proximity to each other (24). This condition is probably not met by any of the crystallins. The absence of formation of Dopa or Bityr suggests that the Tyr residues in the crystallins of the eye lens do not appear to generate any chromophore (or fluorophore) with spectral features in the UV-A and visible region. Trp residues seem to be the only ones capable of doing so.

We need also to mention here that glycation (or nonenzymatic glycosylation) of lysine residues in proteins leads to the production of UV-A absorbing compounds such as pentosidines, furosine and so on. Thus there is an additional chromogenic reaction, namely glycation or the “browning reaction” which is also responsible for the brown colouration of old or cataractous lenses (25–28).

Photodynamic Efficiencies of Endogenous Chromophores

We described above the changes that occur to the lens upon absorption of light in the UV-B region. However, the eye lens does not receive radiation below 310 nm to any significant degree because of the filtering action of the cornea (figure 1). Thus, light induced chemical damage to the lens proteins does not usually arise by the direct absorption process but by the photodynamic action of other molecules which absorb longer wavelengths and cause damage to the protein (substrate) through radical and oxyradical mediated reactions (29–33).

Figure 4 illustrates that such a sensitizer molecule absorbs light in the ground state and eventually reaches the longer lived triplet excited state. From here, it can react with an oxidizable substrate (e.g. amino acids, proteins, lipids) by two major pathways (34). The type I path involves the direct interaction of the triplet sensitizer with the substrate to generate radicals and active oxygen species (H_2O_2 , $\text{O}_2^- \cdot$, $\cdot\text{OH}$), which readily react with the substrate and covalently modify it. In the type II pathway, the sensitizer transfers its energy not to the substrate, but to molecular oxygen and generates the reactive species $^1\text{O}_2$ or singlet oxygen which, in turn, reacts with the substrate and oxidizes it. It may be noticed that in this photodynamic action, the sensitizer absorbs light of wavelength longer than that absorbed by the substrate. In the case of the lens, this means sensitizers that absorb in the UV-A or higher regions; the kynurenine class of chromophores, coenzymes such as the flavins or the NAD class would be the likely candidates (35, 36).

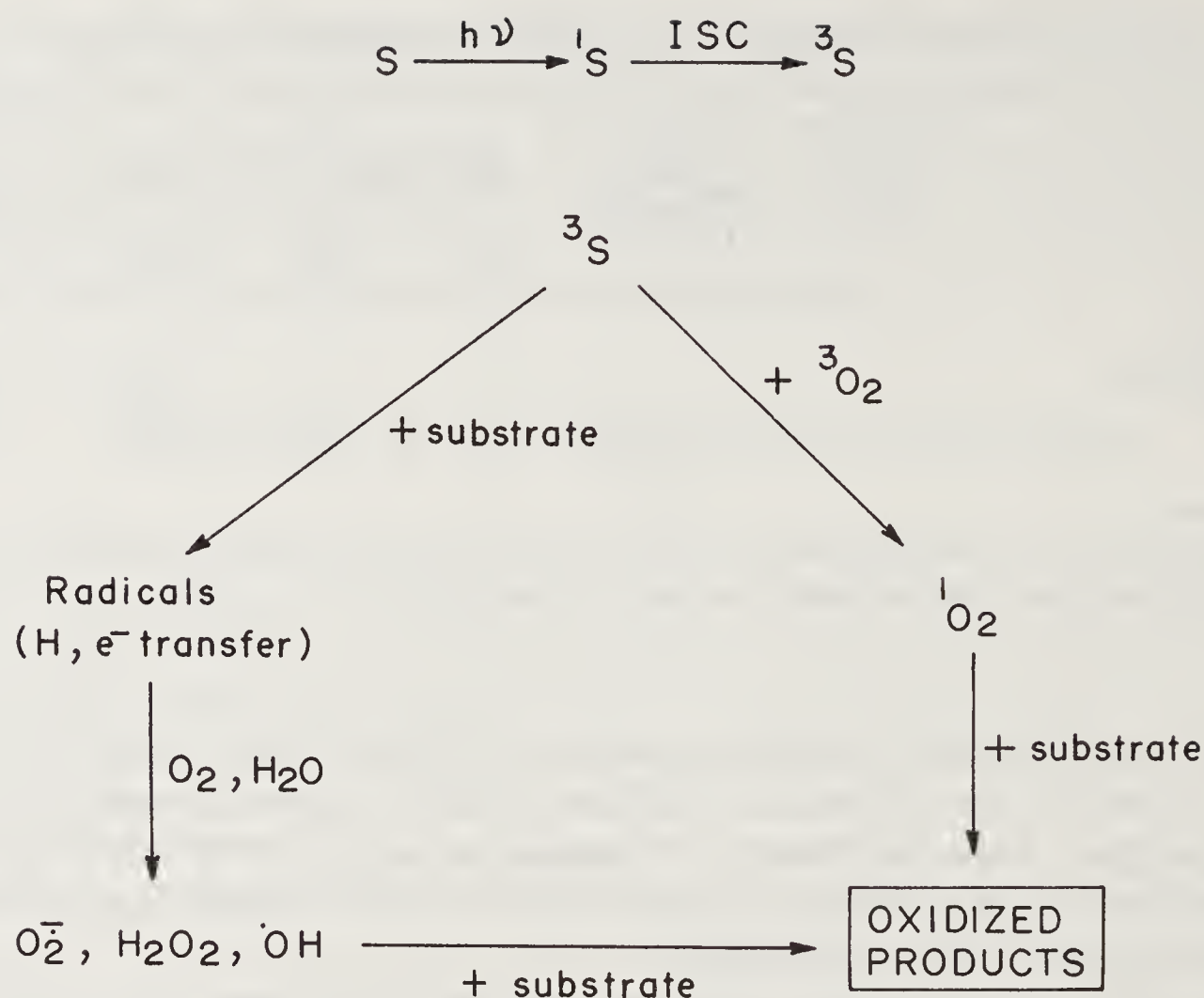


Figure 4. Photodynamic action of a sensitizer (S), 1S and 3S refer to the singlet and triplet excited states of the sensitizer. ISC refers to intersystem crossing.

The question of whether the kynurenines (and other chromophores that accumulate with time in the lens) can mediate photodynamic damage to the lens has been the subject of considerable attention and concern over the years. NFK has been used by several workers for the photodynamic modification of the crystallins (37,38). Flavin induced photodynamic damage has been studied by others (39). However, the first systematic and quantitative studies on the photodynamic efficiencies of several of these compounds has been our own (40). We have monitored the efficiencies of singlet oxygen production (type II pathway) as well as those of superoxide production (type I pathway) of NFK, Ky, KUA, 3HK and the flavins FMN and riboflavin. Singlet oxygen production was monitored by the spectral assay involving the bleaching of the colour of N,N'-dimethyl-4-nitrosoaniline (41), while superoxide production was assayed both by spin trapping ESR spectroscopy (42) and by the ferricytochrome C reduction spectral method (43). The flavins and KUA generate 1O_2 with high efficiency, followed by NFK, and Ky and 3HK generate no singlet oxygen at all. The

Table 2. Photodynamic efficiencies of some endogenous chromophores (40)

Sensitizer	1O_2	$O_2^{\cdot-}$	$O_2^{\cdot-}$ in the presence of NADH (0.1 mM)
Ky	0.00	0.00	0.00
3HK	0.00	0.00	0.00
NFK	0.17	0.01	0.16
KVA	0.42	0.001	0.06
RF, FMN	0.49	0.009	0.225*

*In presence of EDTA, since there appears to be a reaction between NADH and flavins

super-oxide quantum yields of all these compounds are low, and Ky and 3HK produce no detectable amounts of $O_2^{\cdot-}$. The production of $O_2^{\cdot-}$ is enhanced in the presence of electron donors such as NADH. Table 2 highlights these findings in a quantitative fashion, and suggests that kynurenine and 3-hydroxykynurenine are benign components which might act not as sensitizers but as filters that would attenuate light of wavelengths 360 nm and below from the retina.

Oxidative Protein Damage through Photosensitization

We have also looked at the photodynamic efficiencies of the kynurenines (Ky, NFK, 3HK, KUA) towards the oxidative damage and formation of high molecular weight products of δ -crystallin, the core protein of chicken eye lens. Figure 5 compares the changes that occur to the molecular size of δ -crystallin upon direct photolysis at 300 nm, as well as upon irradiation in the presence of various kynurenine derivatives, with the wavelengths of irradiation corresponding to the appropriate absorption band maxima. The electrophoretogram displayed in figure 5a was run after the addition of sodium dodecylsulphate (so as to keep the molecules in the disaggregated and monomeric form) and 2-mercaptoethanol (so as to monitor only non-disulphide covalent crosslinking). Direct photolysis of the protein is the most effective in causing covalent damage. Not only are higher molecular weight products seen (bands above

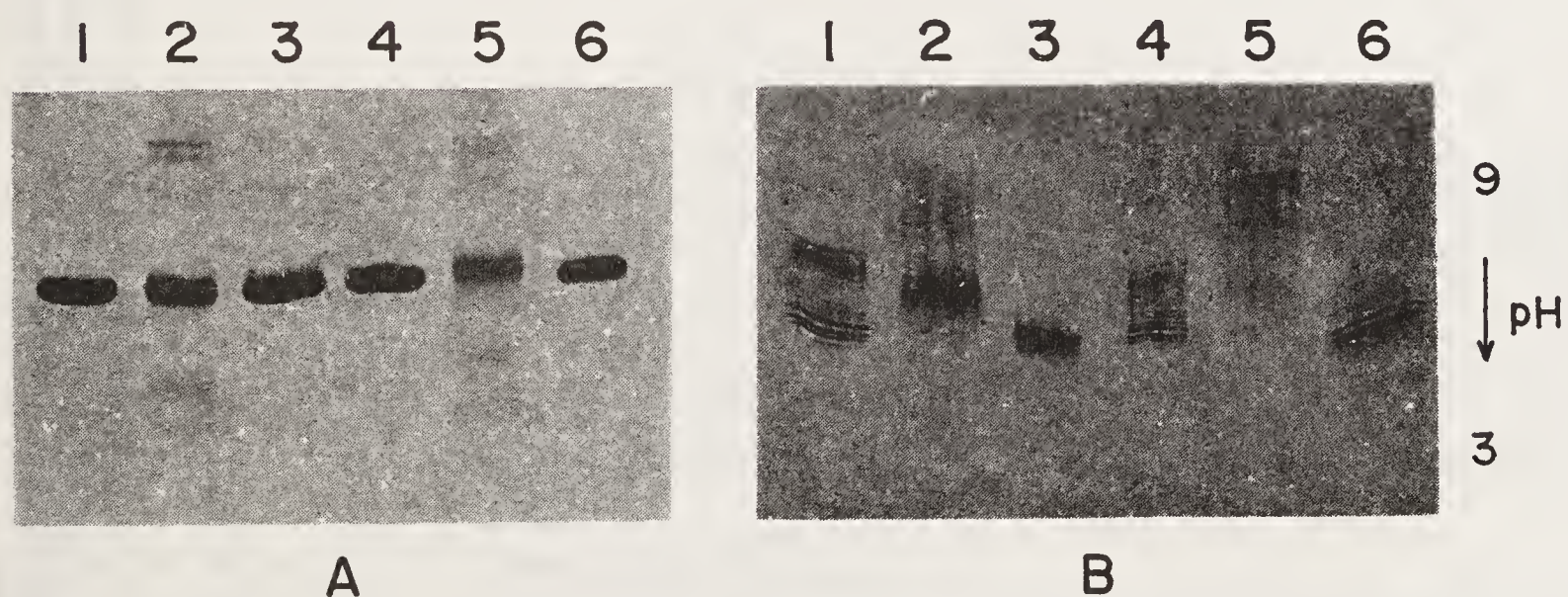


Figure 5. Polyacrylamide gel electrophoregrams (PAGE) of photodynamically damaged δ -crystallin. A solution of δ -crystallin (0.6 mg/ml) was irradiated for 60 minutes in the sample compartment of a Hitachi F-4000 spectrofluorimeter, in the presence of various photosensitizers, with the excitation monochromator set at wavelengths corresponding to the absorption maxima of the photosensitizers. The excitation bandpass was set at 20 nm. Panel A shows a 12.5% SDS-PAGE on which samples were electrophoresed after they were boiled with sample buffer containing sodium dodecylsulphate (SDS) and β -mercaptoethanol. Panel B shows an isoelectric focussing (IEF) gel of pH range 3–9. In both panels, lane 1 corresponds to control protein, lane 2 to protein photodamaged with light of wavelength 295 nm, lane 3 to protein irradiated with 320 nm light in the presence of 2 mM NFK, lane 4 to protein irradiated with 360 nm light in the presence of 2 mM Kyn, lane 5 to protein irradiated with 332 nm light in the presence of 2 mM KUA and lane 6 to protein irradiated with 360 nm light in the presence of 2 mM 3HK.

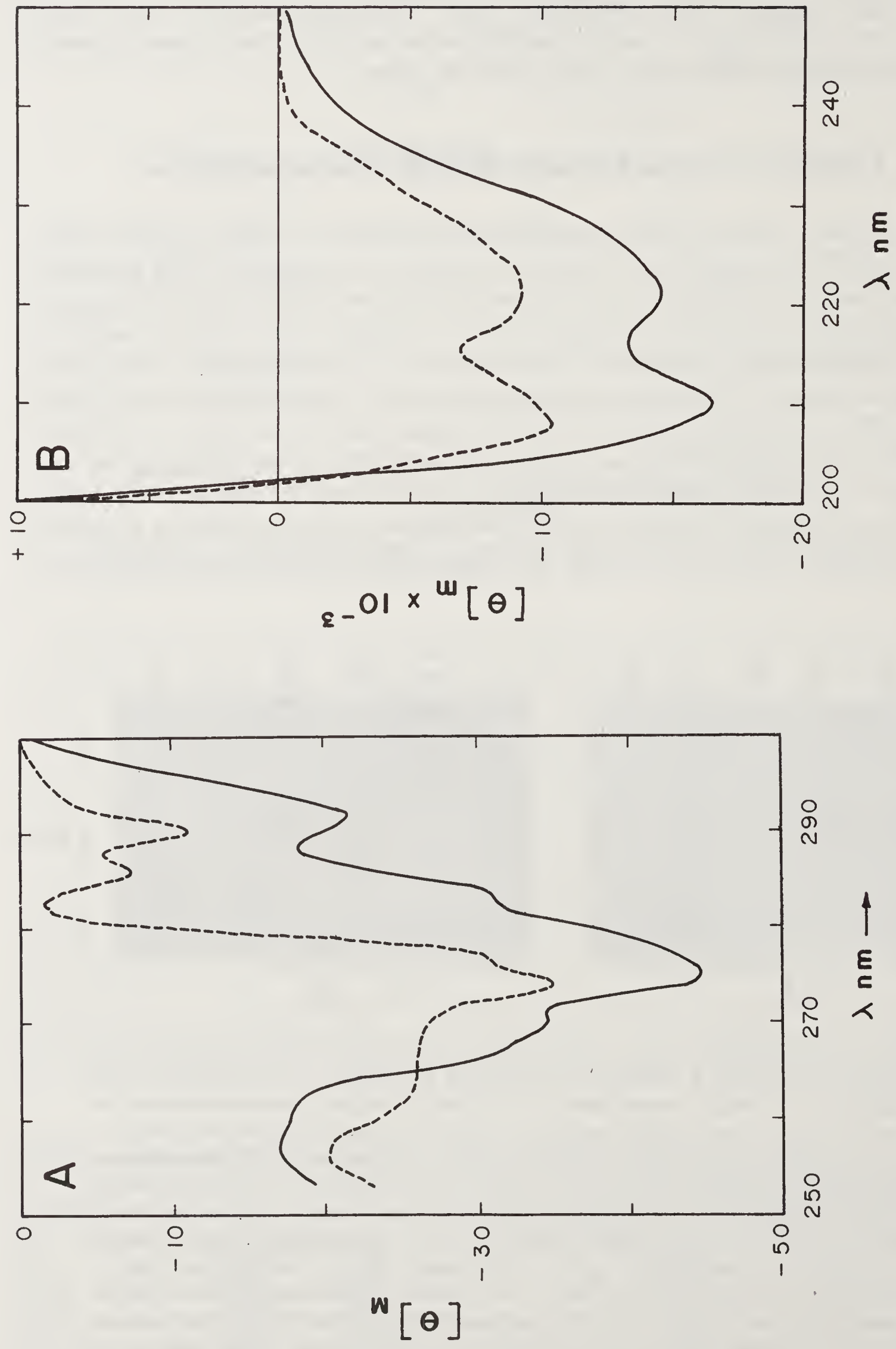


Figure 6. Panels A and B show the near and far UV circular dichroic (CD) spectra, respectively, of δ -crystallin and δ -crystallin irradiated for 60 minutes with light of wavelength 360 nm (and bandpass 20 nm) in the presence of 5mM KUA, in the sample compartment of a Hitachi F-4000 spectrofluorimeter. The CD spectra were recorded on a Jasco J-20 spectropolarimeter.

the parent crystallin band), but also some lower molecular weight ones (seen below the control band), indicative of photolytic covalent crosslinking as well as degradation or proteolysis. Sensitization by NFK does occur (lane 3) to produce a feeble high molecular weight protein, but no degradation products are detectable, at least in this experiment. KUA is clearly a better sensitizer as seen in lane 5, and photodynamically produces both covalent polymers and degradation products of δ -crystallin. Kynurenine and 3HK do not generate any photodynamic modifications in the protein, as lanes 4 and 6 reveal. Figure 5b shows the isoelectric focussing (IEF) electrophoretogram of the same samples. Lanes 2 and 5, corresponding to photodamaged and KUA photosensitized protein, show that charge differences leading to changes in the pI of the protein, occur to a significant extent in both cases, while lanes 4 and 6 corresponding to Ky and 3HK do not show such changes. This behavior of Ky and 3HK is consistent with the results of table 2 and support the suggestion that they act not as sensitizers but as filters of radiation below 360 nm. Dillon and Atherton (44) have independently come to the same conclusion about 3HK and 3HKG, based on their very low emission quantum yields and small lifetimes in the excited state.

Figure 6 shows the circular dichroic spectra in the near and far UV regions of δ -crystallin damaged through the photosensitization of KUA, in comparison with that of undamaged protein. As can be seen from the spectra, changes occur in both the tertiary structure (around the aromatic sidechains at least) and the secondary structure of the protein upon photosensitized damage.

Conclusion

The transparency of the eye lens is intimately related to the state of organization of the fibre cells and their constituent crystallins. Disruption of protein packing may result in discontinuities in the refractive index gradient at the very least, and in the precipitation of protein in more severe instances. Both these phenomena would adversely affect the transmission of light through the lens to the retina.

Alterations in the packing of a heterogeneous set of protein molecules may arise from

- (i) Alterations in the charge profile of the protein,
- (ii) Conformational changes at the secondary and tertiary structural levels, affecting the intermolecular interactions and packing, and
- (iii) Changes in protein size resulting from the formation of covalent crosslinks that could change the hydropathy of the protein, and cause its precipitation.

Of the twenty amino acid residues in proteins, the ones that possess reactive side chain groups which can lead to protein modification or crosslinking are Cys, Met, His, Trp, Tyr, Lys, Arg, Glu, Asp and Ser. Amongst these, Cys would lead directly to disulphide crosslinks, while amide or isopeptide bonds are possible between Glu (or Asp) and Lys (or Arg). Met would be oxidised to its sulfoxide or sulphone. The oxidation of His, Trp and Tyr are of interest because of chromophore alterations and potential non-disulphide crosslinks. We have investigated (18, 19, 45) the damage caused to proteins in general, and crystallins in particular, by oxyradical species such as 1O_2 and $\cdot OH$, and have shown that they cause changes in charge and structure, as well as non-disulphide and non-dityrosine interprotein crosslinks. Our data suggest

that these crosslinks involve the oxidation products of the imidazole ring of the His residues on the one hand, and the free amino groups of lysine residues, and might occur through such sidechain isopeptide or amide bonds.

Modification of the Trp and Tyr residues leads to chromophores that absorb in the UV-A or even the visible region. We have investigated the photodynamic efficiencies of several of these, as well as some others which are endogenous to the eye lens (40). Some of these molecules can produce oxyradicals upon irradiation with visible light, which is incident on the eye lens much of the time. In addition, we are also investigating the role that glycation products may play in sensitization, and in affecting changes in the charge and structural features of the crystallins.

Acknowledgement

We are grateful to Mrs. Shaila and Dr. Gullapalli N. Rao of the L. V. Prasad Eye Institute, Hyderabad, for gifts of normal and cataractous human lenses. We thank S. Chenchal Rao for help and advice.

References

1. Bloemendal, H. (1981) *Molecular and Cellular Biology of the Eye Lens* (John Wiley & Sons, New York).
2. Lerman, S. (1980) *Radiant Energy and the Eye* (Macmillan, New York).
3. Tardieu, A & Delaye, M. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17**, 47–70.
4. Sliney, D. H. (1986) *Invest. Ophthalmol. Vis. Sci.* **27**, 781–790.
5. Zigman, S. & Paxhia, T. (1988) *Exp. Eye. Res.* **47**, 815–824.
6. Rao, Ch.M., Balasubramanian, D. & Chakrabarti, B. (1987) *Photochem. Photobiol.* **46**, 511–515.
7. van Heyningen, R. (1973) in *CIBA Symp. (New Series)*, (Elsevier, Amsterdam, Holland), pp. 151–164.
8. Pirie, A. (1971) *Biochem. J.* **125**, 203–211.
9. Waxler, M. & Hitchins, V. M. (1986) *Optical Radiation and Visual Health* (CRC Press, Boca Raton, FL, USA).
10. Truscott, R. J. W., Faull, K. & Augusteyn, R. C. (1977) *Ophthalmic Res.* **9**, 263–268.
11. van Haard, P. M., Hoenders, H. J. & Ketelaars, H. C. J. (1980) *Ophthalmic Res.* **12**, 252–255.
12. Zigman, S., Datiles, M. & Torczynski, E. (1979) *Invest. Ophthalmol. Vis. Sci.* **18**, 462–469.
13. Jamieson, S. W. (1970) *Postgrad. Med. J.* **46**, 557–559.
14. Frederick, J. E., Snell, H. E. & Haywood, E. K. (1989) *Photochem. Photobiol.* **50**, 443–450.
15. Taylor, H. R. (1989) *Photochem. Photobiol.* **50**, 489–492.
16. Saito, I., Matsuura, T., Nagakawa, M. & Hino, T. (1977) *Acc. Chem. Res.* **10**, 346–352.
17. Wasserman, H. H. (1970) *Ann. N.Y. Acad. Sci.* **171**, 108–120.
18. Balasubramanian, D., Du, X. Y. & Zigler, J. S., Jr. (1990) *Photochem. Photobiol.* **52**, 761–768.
19. Guptasarma, P. & Balasubramanian, D., submitted.
20. McNamara, M. & Augusteyn, R. (1984) *Exp. Eye. Res.* **38**, 45–56.
21. Joshi, P. C., Carraro, C. & Pathak, M. A. (1987) *Biochem. Biophys. Res. Commun.* **142**, 265–274.
22. Amado, R., Aeschbach, R. & Neukom, H. (1984) *Methods Enzymol.* **107**, 377–388.
23. Garcia-Castineiras, S., Dillon, J. & Spector, A. (1978) *Exp. Eye Res.* **26**, 461–476.
24. Anderson, S. R. & Malencik, D. A. (1989) in *Fluorescent Biomolecules*, eds. Jameson, D. M. & Rein, G. D. (Plenum Press, New York), pp. 217–245.
25. Monnier, V. M. & Cerami, A. (1981) *Science* **211**, 891–893.
26. Njoroge, F. G., Fernandes, A. A. & Monnier, V. M. (1988) *J. Biol. Chem.* **263**, 10646–10652.

27. Abraham, E. C., Swamy, M. S. & Perry, R. E. (1988) *Prog. Clin. Biol. Res.* **304**, 123–139.
28. Ahmed, M. U., Dunn, J. A., Walla, M. D., Thorpe, S. R. & Baynes, J. W. (1988) *J. Biol. Chem.* **263**, 8816–8821.
29. Zigler, J. S., Jr. & Goosey, J. D. (1981) *Photochem. Photobiol.* **33**, 869–874.
30. Zigler, J. S., Jr., Jernigan, H. M., Jr., Perlmutter, N. S. & Kinoshita, J. (1982) *Exp. Eye Res.* **35**, 239–249.
31. Zigman, S. (1981) in *Mechanisms of Cataract Formation in the Human Lens*, ed. Duncan, G. (Academic Press, New York), pp. 117–150.
32. Chakrabarti, B., Bose, S. K. & Mandal, K. (1986) *J. Indian Chem. Soc.* **63**, 131–137.
33. Jori, G. & Spikes, J. D. (1981) in *Oxygen and Oxyradicals in Chemistry and Biology*, ed. Rodgers, M. A. J. & Powers, E. L. (Academic Press, New York), pp. 441–460.
34. Foote, C. S. (1968) *Acc. Chem. Res.* **1**, 104–110.
35. Goosey, J. D., Zigler, J. S., Jr. & Kinoshita, J. (1980) *Science* **208**, 1278–1280.
36. Dillon, J. (1984) *Curr. Eye. Res.* **3**, 145–150.
37. Mandal, K., Bose, S. K. & Chakrabarti, B. (1986) *Photochem. Photobiol.* **43**, 515–523.
38. Walker, M. L. & Borkman, R. F. (1989) *Exp. Eye Res.* **48**, 375–383.
39. Bose, S. K., Mandal, K. Chakrabarti, B. (1986) *Photochem. Photobiol.* **43**, 525–528.
40. Muralikrishna, C., Uppuluri, S., Riesz, P., Zigler, J. S., Jr. & Balasubramanian, D. (1990) *Photochem. Photobiol.*, submitted.
41. Kraljic, I. & El Mohsni, S. (1979) *Photochem. Photobiol.* **28**, 577–581.
42. Finkelstein, E., Rosen G. M. & Rauckman, E. J. (1982) *Mol. Pharmacol.* **21**, 262–265.
43. McCord, J. M. & Fridovich, I. (1969) *J. Biol. Chem.* **244**, 6049–6055.
44. Dillon, J. & Atherton, S. J. (1990) *Photochem. Photobiol.* **51**, 465–468.
45. Balasubramanian, D. (1991) in *Biological Oxidation*, ed. Hamilton, G. A., Reddy, C. C. & Madyastha, K. M. (Academic Press, New York), Vol. 2, pp. 953–964.

Molecular Structures of Spermine: Nucleic Acid Complexes

Loren Dean Williams, Christin A. Frederick, Reinhard V. Gessner and Alexander Rich

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA 02139

Introduction

The polyamines, putrescine, spermidine and spermine, are ubiquitous within all types of cells ranging from simple procaryotes to complex eucaryotes (for reviews see references 1–3). Polyamines are critical components of biological systems and are involved in a diverse array of primary processes. In the nuclei of eucaryotic cells, the concentrations of spermine and spermidine are of the order of one millimolar while the concentration of putrescine is ten micromolar. Biosynthesis of polyamines is regulated and the concentrations increase when cell growth is stimulated; conversely, deprivation of polyamines inhibits cell division.

Polyamines interact directly with anionic cellular components such as membranes (for a recent review see reference 4) and nucleic acids. Polyamines are known to stabilize duplex DNA against thermal denaturation (5–8), can condense DNA (9–12) and chromatin (13, 14) into compact structures, and promote the B-DNA to Z-DNA transition (15–17). Polyamines increase the rate of movement of DNA replication forks (18), and the fidelity of restriction endonucleases (19). They also interfere with DNA binding of antibiotics such as actinomycin-D (20). With affinity probes it has been shown that polyamines recognize discrete binding sites on nucleosome core particles (21), as well as on yeast-tRNA^{asp} and its synthetase complex (22).

Although the importance of polyamines in nucleic acid structure and biology is well recognized, many of the molecular aspects of polyamine:nucleic acid interactions remain to be unraveled. One area of uncertainty centers on specificity of polyamine binding. Polyamines may bind to nucleic acids by charge–charge interactions to form delocalized condensates. This mode of association is driven by release of other bound ions (7, 8, 23, 24) as formulated in the theory of Manning (25). If polyamines bind to DNA in this manner, the complexes should be highly dynamic and lack significant structural and sequence specificity. Alternatively, polyamines may bind to nucleic acids by additional, more direct molecular interactions (26–32), such as hydrogen bonds and hydrophobic interactions. In this case, structural features would contribute to stability of polyamine:nucleic acid complexes. Such complexes might then be relatively static and display nucleic acid structural and sequence specificity. A second area of uncertainty centers on the relationships between polyamine binding and nucleic acid conformation and tertiary interactions. Polyamines may modulate the

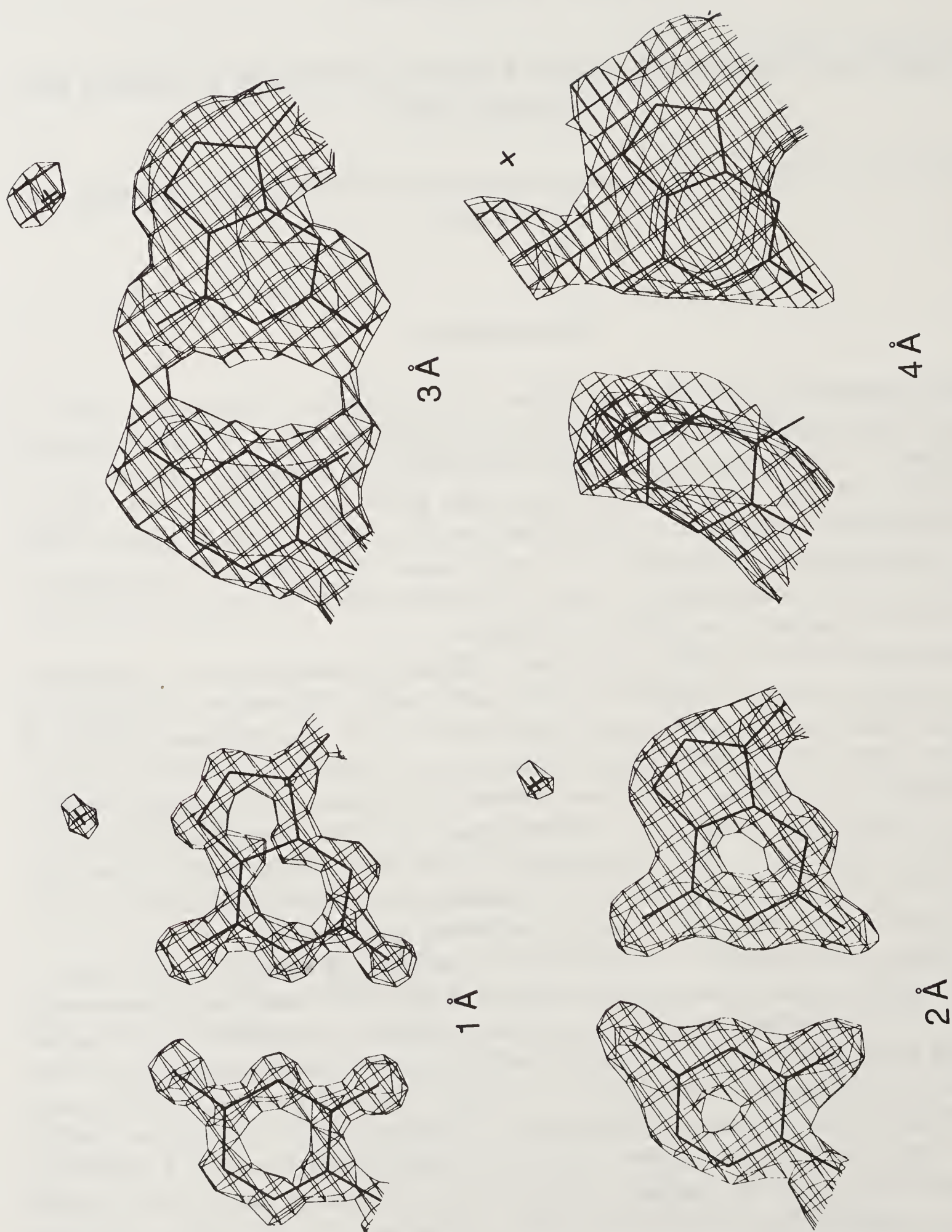


Figure 1. The effect of resolution on an electron density map surrounding a DNA base pair. The results are shown at 4.0 Å; 3.0 Å; 2.0 Å; and 1.0 Å. The X represents a water molecule. The base pair is enveloped by a net representing electron density.

stability of specific nucleic acid conformations or nucleic acid:drug and nucleic acid:protein complexes. Alternatively monovalent and divalent monocations, albeit in greater concentrations, may effectively substitute for polyamines in modulating stability of 'high salt' nucleic acid conformations and complexes. In an attempt to clarify and address some of the uncertainties surrounding the role of polyamines in nucleic acid structure, we review here the three-dimensional structures of spermine:nucleic acid complexes as determined by X-ray crystallography. Crystal structures of other polyamines bound to nucleic acids have not yet been determined.

The information available from a crystal structure is determined largely by the degree of order within the crystal. The crystalline order establishes the resolution of the diffraction data and dictates the degree of detail in the electron density maps used to build and refine molecular models. An example of the effects of resolution on electron density maps is shown in figure 1. With resolution of 4.0 Å, only rudimentary forms are observable and although it may be possible to trace a macromolecular backbone, sidechains will remain obscure. With resolution of 3.0 Å, amino acid and nucleic acid sidechains are discernible. With resolution of 2.0 Å, many solvent molecules can be located with confidence in the map. With resolution of 1.0 Å, (atomic resolution) each atom forms a discrete ball of electron density. The resolution of the spermine:nucleic acid complexes described here varies. In certain instances the resolution severely limits the amount of information.

The simplest polyamine, putrescine, contains two positive charges at physiological pH. It is composed of four methylene groups and two amino groups. Putrescine is a biological precursor of spermidine, the intermediate polyamine. Spermidine, with three positive charges, is formed by addition of three methylenes and one amino group to putrescine and thus contains seven methylenes and three amino groups. Spermine, with four positive charges, is formed by addition of three methylenes and one amino group to spermidine and thus contains ten methylenes and four amino groups. The structure of the phosphate salt of spermine was solved by X-ray crystallography (33). The bond distances and dihedral angles of spermine within this crystal are shown in figure 2. In addition, spermine has been cocrystallized with four classes of nucleic acids, tRNA (34), A-DNA (35), B-DNA:anthracycline complexes (32, 36) and Z-DNA (37). The complex formed by spermine bound to yeast-tRNA^{Phe} diffracted to 2.7 Å, while the complex with A-DNA diffracted to 2.0 Å, the B-DNA:anthracycline complexes diffracted to around 1.5 Å and the complex with Z-DNA diffracted to 1.0 Å. A spermine molecule was tentatively identified bound to a B-DNA dodecamer (38).

However, this putative spermine molecule was never fully refined and will not be discussed further here.

Yeast-tRNA^{Phe}

The addition of polyamines to yeast-tRNA^{Phe} is required to produce crystals that diffract to reasonable resolution (2.7 Å) (39, 40). This might be expected from the critical role of polyamines in stabilizing tRNA conformation (41). In the final refinement of the yeast-tRNA^{Phe} structure, two spermine molecules and four hydrated magnesium atoms were located in the electron density maps (34). Both spermine molecules bind in the grooves of tRNA. The first of these spermine molecules is found in the deep groove (figure 3A), extending from one end of the D stem into the anticodon stem. Although detailed assignments of intermolecular interactions remain uncertain at

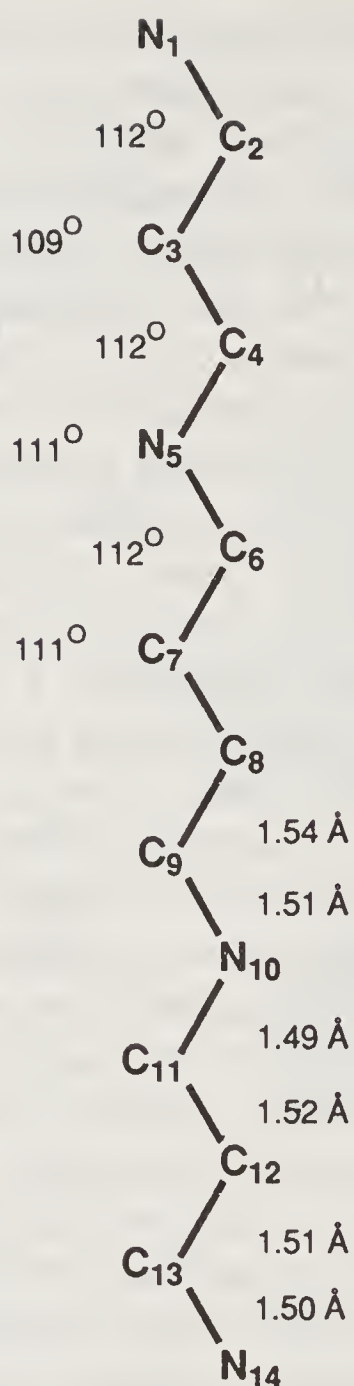


Figure 2. Bond angles and distances from the X-ray crystal structure of spermine:phosphate (33).

2.7 Å resolution, three of the amino groups of this spermine molecule are near the phosphate oxygens of yeast-tRNA^{Phe} and may form hydrogen bonds to these oxygens. In addition, one central amino group of this spermine molecule appears to be within hydrogen-bonding distance of the O6 of guanine. The second spermine molecule bound to yeast-tRNA^{Phe} is located in the minor groove of the D stem near the intersection of the acceptor stem (figure 3B). This spermine molecule, like the first, is near several phosphate oxygens, and may form hydrogen bonds with them. In addition, this spermine molecule is in a position which suggests that it could be forming hydrogen bonds to several ribose oxygens. However, it should be emphasized that trying to fix the detailed interactions of spermine molecules at 2.7 Å resolution is difficult and the results should be regarded as tentative. It will be interesting to see if higher resolution studies of this crystal form are capable of yielding further information about spermine interactions.

A-DNA

In the A-DNA helix, the major groove is relatively deep and narrow while the minor groove is relatively shallow and wide (42). The structure of spermine bound to the

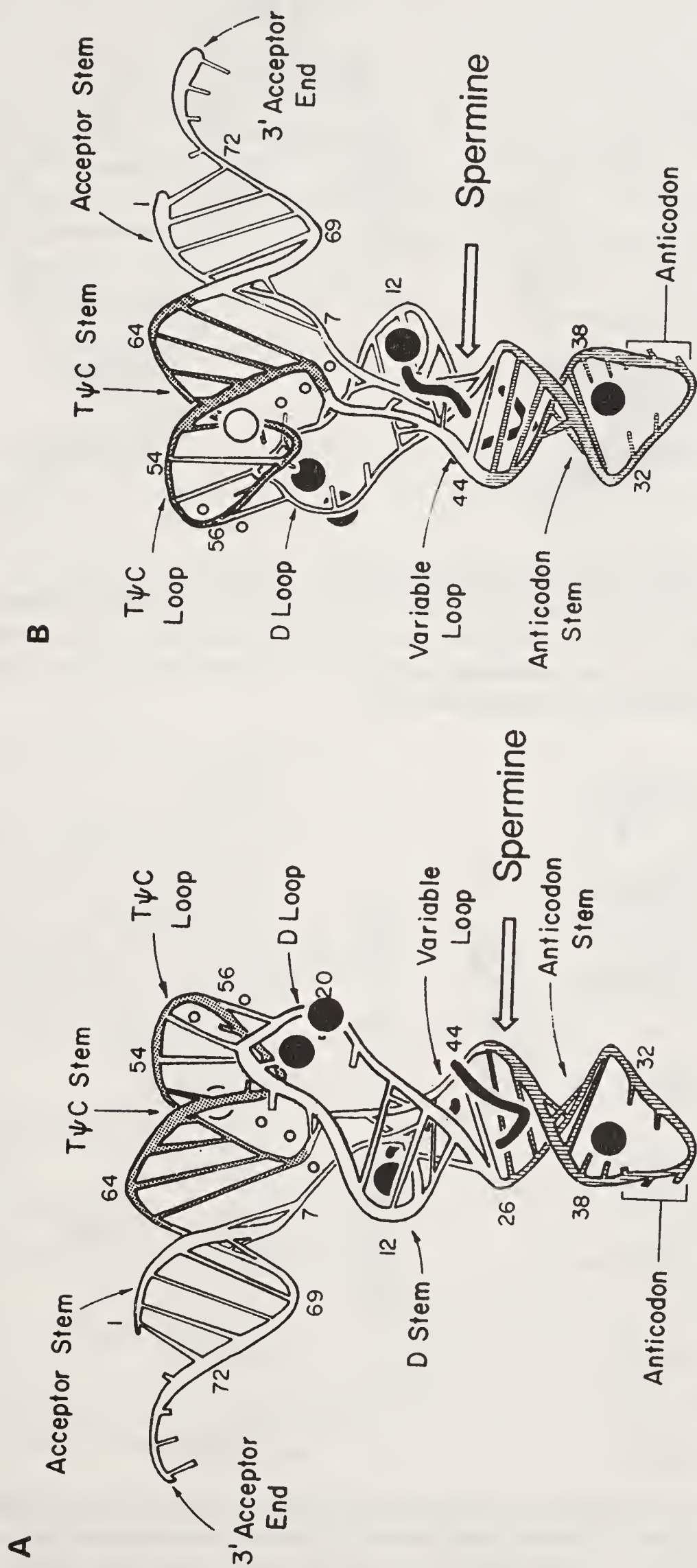


Figure 3. Two sides of the yeast-tRNA^{Phe} molecule. The hollow tube represents the tRNA backbone and the crossbars represent the base pairs. The spermine molecules are shown in heavy black lines and are indicated by arrows. Magnesium atoms are solid circles and a tentative magnesium atom is the hollow circle. [Modified from reference (34)].

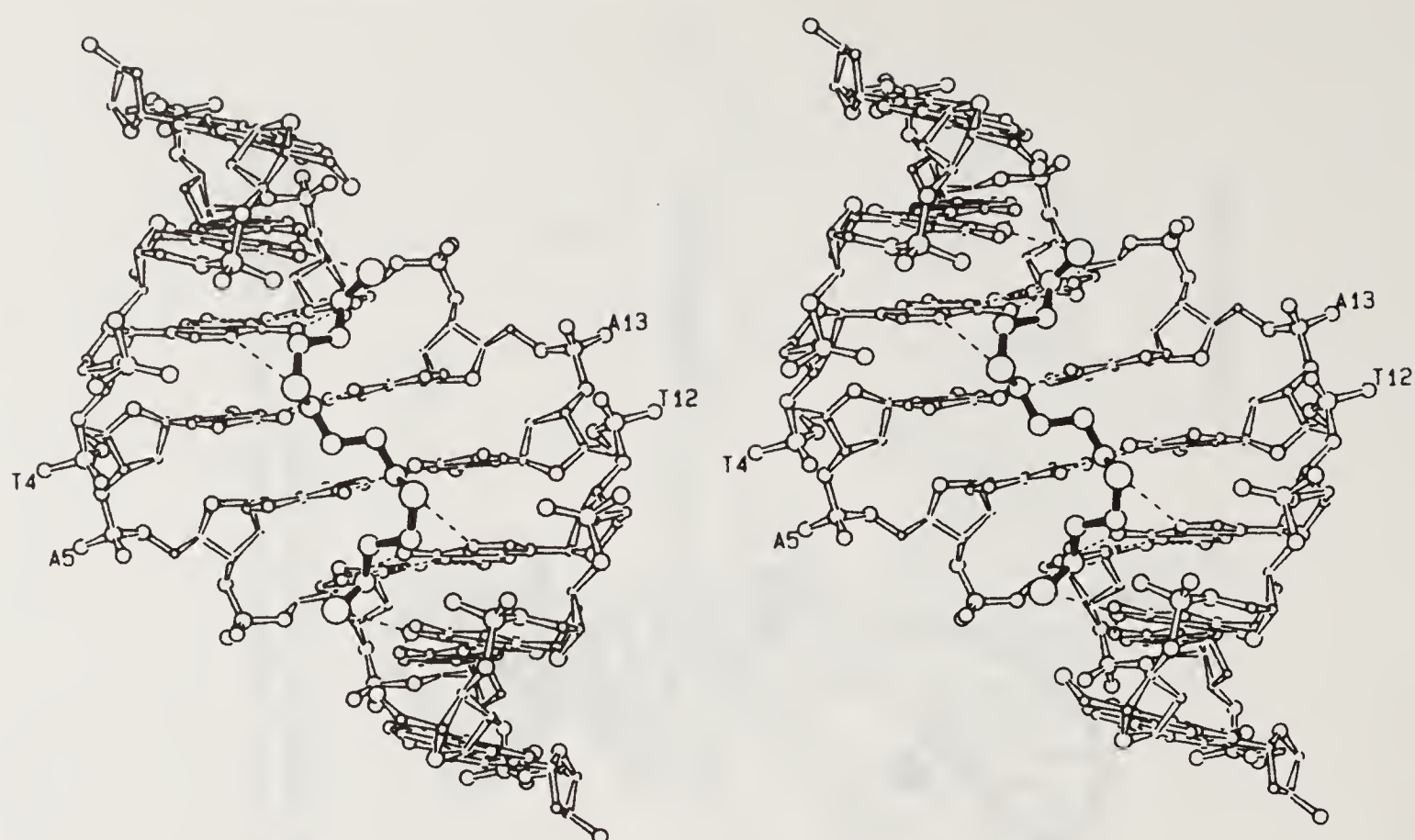


Figure 4. Stereo picture showing spermine bound in the major groove of an A-DNA octamer. The DNA is drawn with hollow bonds and the spermine molecule is drawn with solid bonds. The nitrogen atoms are the largest spheres. Hydrogen bonds between the DNA and spermine molecule are drawn with dashed lines. [Reprinted with permission from reference (35)].

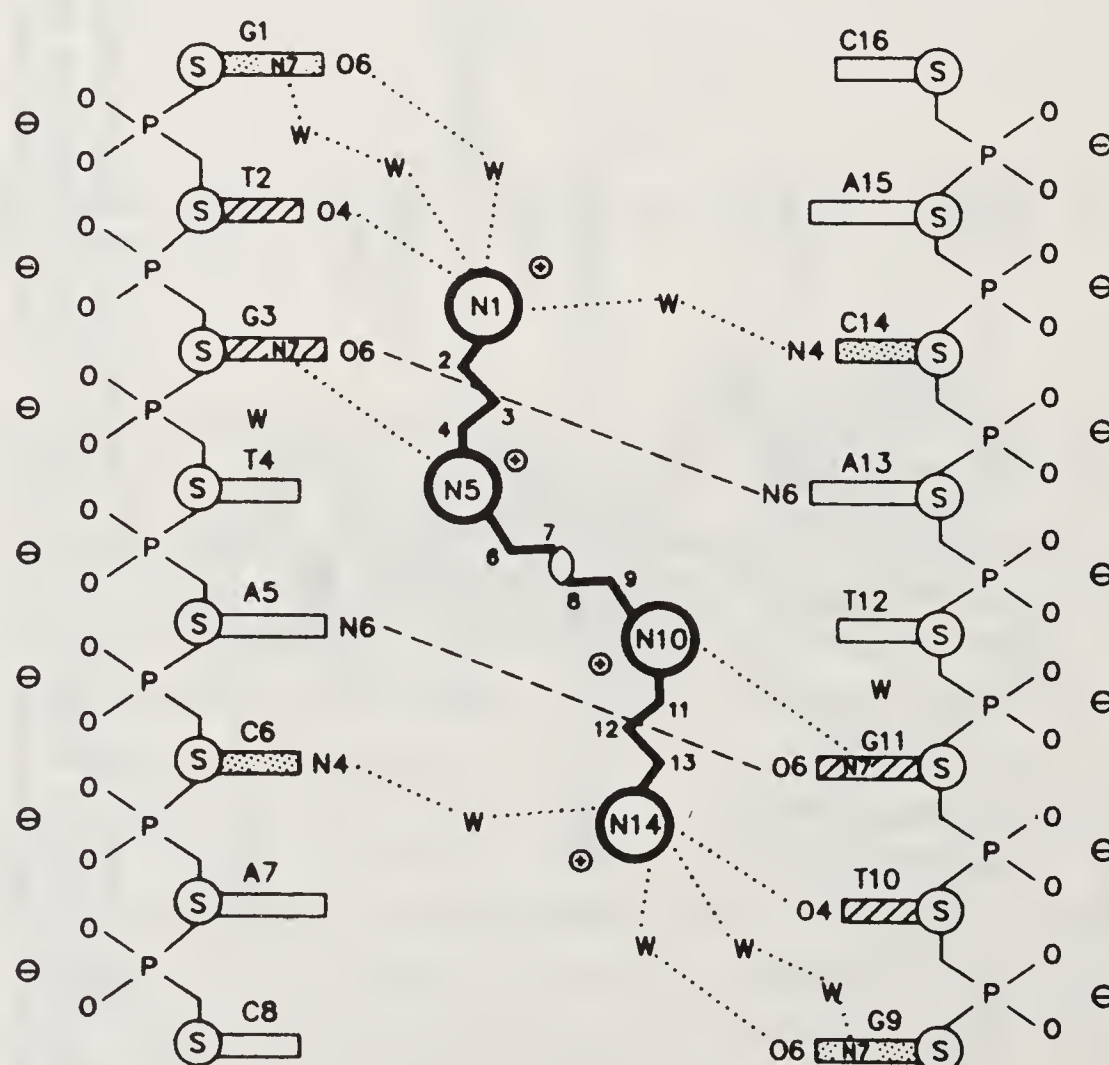


Figure 5. Schematic diagram of spermine molecules bound in the major groove of an A-DNA octamer. The striped and dotted bases are involved in direct and water-mediated hydrogen bonds with the spermine molecule, respectively. The dashed lines represent some of the hydrogen bonds between the two strands of the DNA. The dotted lines represent hydrogen bonds between the spermine molecule and the DNA. The van der Waals contacts are not shown. [Reprinted with permission from reference (35)].

major groove of an A-DNA octamer was determined at 2.0 Å resolution (35). The DNA duplex in this complex is formed from the self-complementary oligonucleotide d(GTGTACAC). A spermine molecule binds to the floor of the major groove (figure 4) and is located on a twofold axis. The amino groups of the spermine molecule form hydrogen bonds to the DNA bases but not to the phosphate oxygens. Each terminal amino group forms a hydrogen bond to the O4 of a thymine and each central amino group forms a hydrogen bond to the N7 of a guanine (figure 5). The spermine molecule does not form hydrogen bonds to phosphate oxygens. However, the spermine molecule does form extensive hydrophobic contacts with the DNA. The most significant of these contacts appear to involve the central methylenes of spermine and the 5-methyl groups of thymines.

DNA:Anthracycline Complexes

Anthracyclines are DNA intercalators which constitute a widely used family of chemotherapeutic agents with a strong dependence of biological activity on minor modifications in chemical structure (43). Daunomycin, adriamycin and 4'-epiadriamycin have each been cocrystallized with the DNA hexamer d(CGATCG) and solved to near 1.5 Å resolution (32, 36). Two drug molecules intercalate at the d(CG)

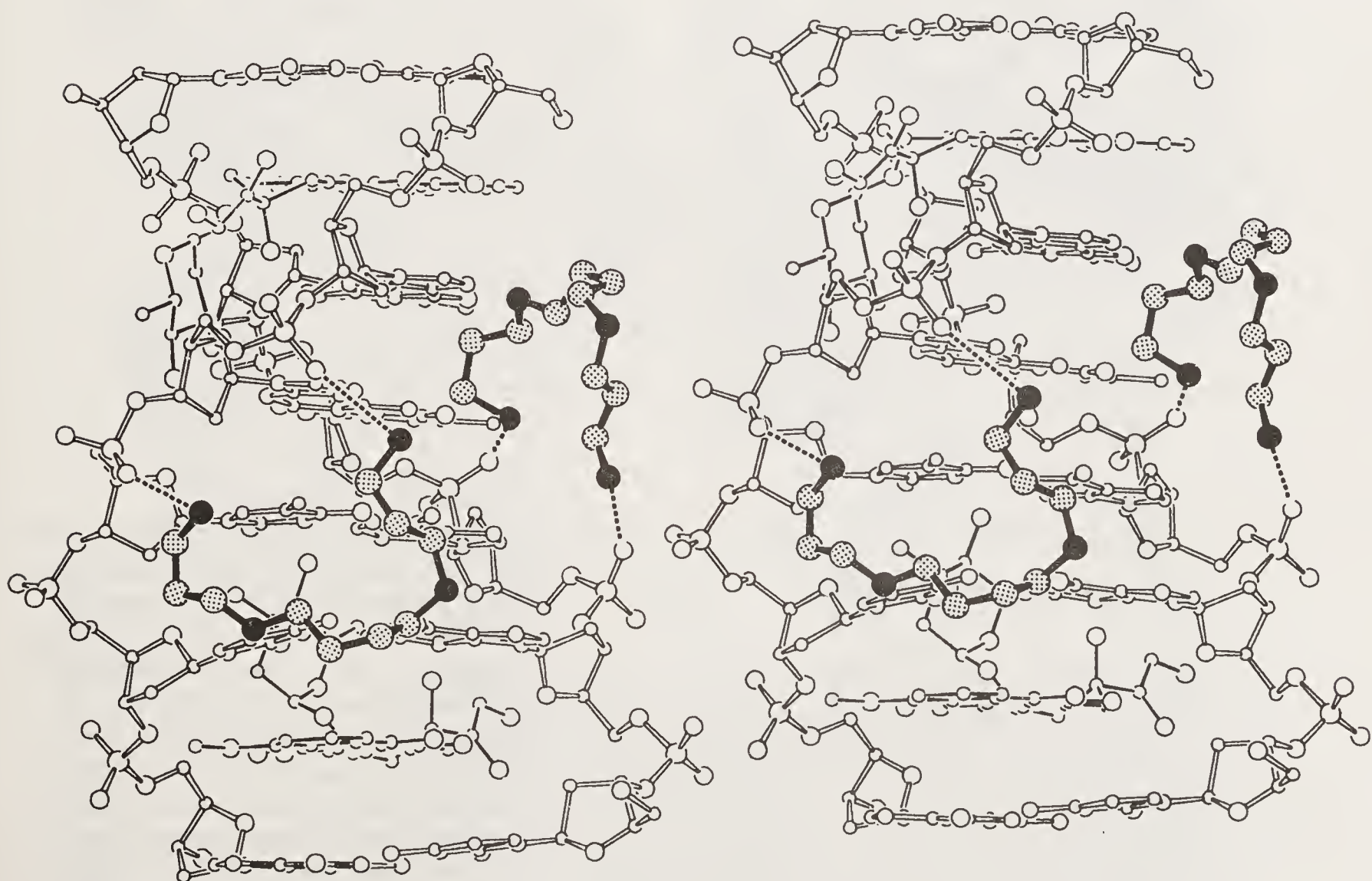


Figure 6. Stereoview ORTEP representation of spermine molecules bound in the major groove of the d(CGATCG):4'-epiadriamycin complex. The DNA is drawn with hollow bonds and spermine is drawn with thick solid bonds. A twofold axis in the center of the DNA duplex relates the two spermine molecules to each other. The nitrogen atoms of spermine are black and the carbon atoms are stippled. The hydrogen bonds linking the spermine with the DNA are drawn with dashed lines. [Reprinted from reference (32)].

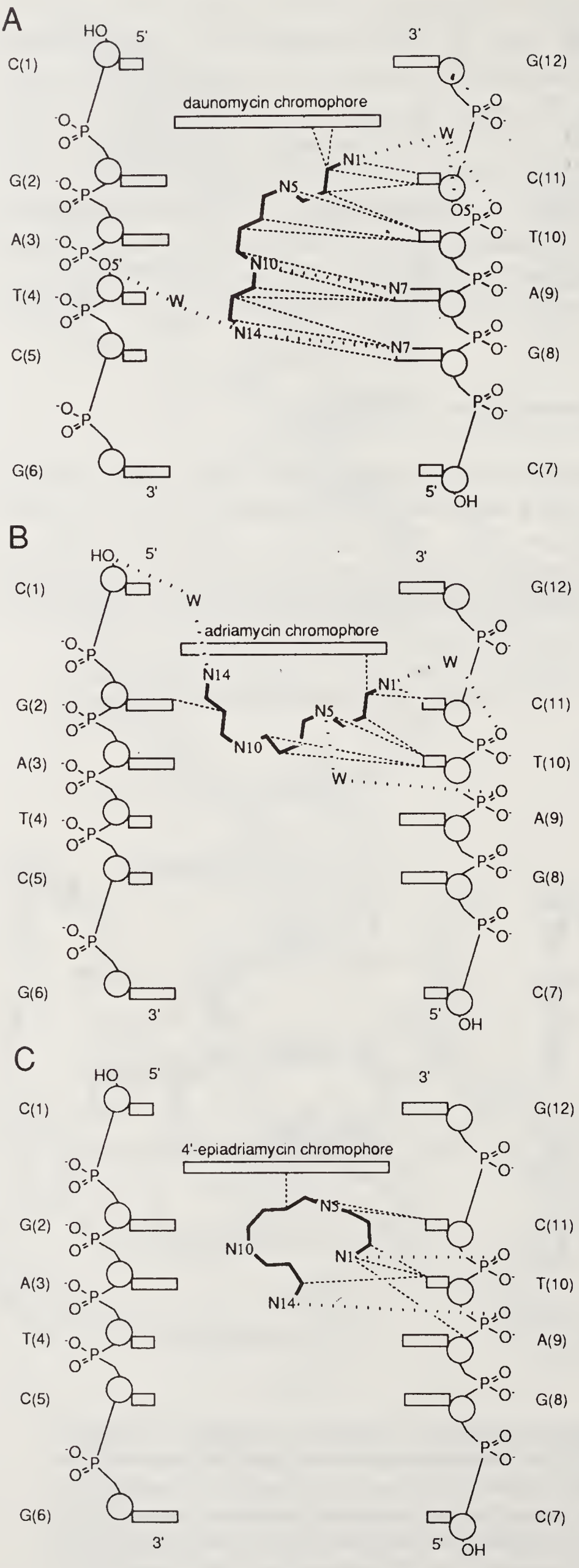


Figure 7. Schematic diagrams of the spermine molecules bound to DNA:anthracycline complexes. (A) The daunomycin:d(CGATCG) complex, (B) The adriamycin:d(CGATCG) complex, (C) The 4'-epiadriamycin:d(CGATCG) complex. Hydrogen bonds are drawn with dashed lines and van der Waals contacts with dotted lines. Atoms of the DNA that form hydrogen bonds to spermine molecules are identified by atom type. For clarity, van der Waals contacts between the spermine molecule and the backbone atoms are omitted. [Reprinted from reference (32)].

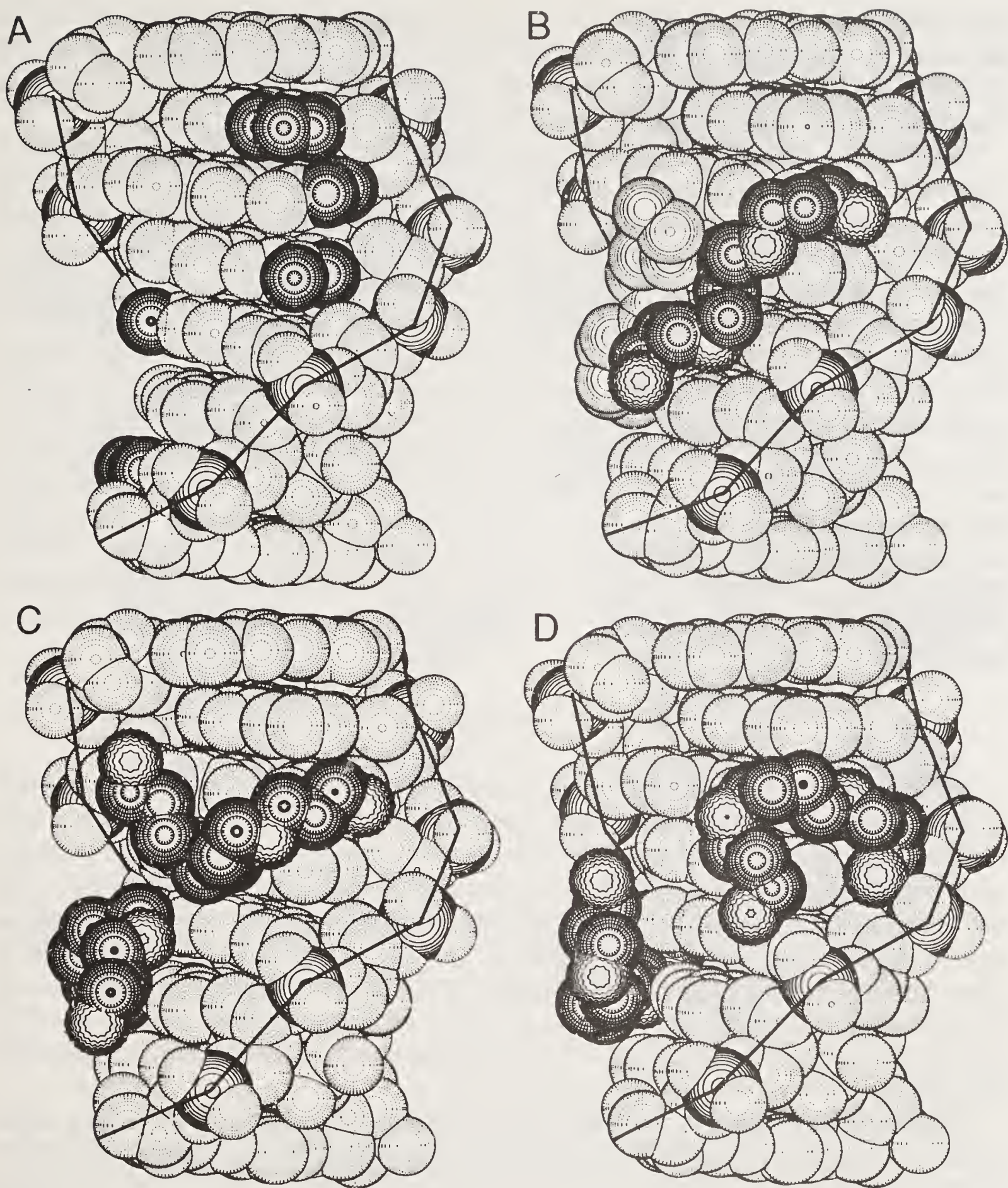


Figure 8. The hydrophobic zone and three spermine:DNA:anthracycline complexes. Space filling representation of the major groove of (A) The d(CGATCG):4'-epiadriamycin complex, minus the spermine molecules, but with the hydrophobic zone highlighted, (B) The spermine:d(CGATCG):daunomycin complex, (C) The spermine:d(CGATCG):adriamycin complex, and (D) The spermine:d(CGATCG):4'-epiadriamycin complex. Phosphorous atoms are marked with solid circles. In (A) the hydrophobic zone and in (B–D) the spermine carbon atoms are marked with circles with radial lines. The spermine nitrogen atoms in (B–D) are marked with wavy circles. For clarity the spermine molecule on the left side of (B) is marked by lighter, dotted circles. The DNA backbones are traced with solid lines. [Reprinted from reference (32)].

steps of each hexamer duplex and two symmetry-related spermine molecules bind in the major groove. The only covalent differences between the complexes are in the hydroxyl groups of the anthracyclines. The spermine:d(CGATCG):4'-epiadriamycin complex is shown as an example of this type of complex in figure 6.

The conformation of the DNA is nearly the same in each of these complexes. However, the solvent structure is modified apparently by the changes in the hydroxyl groups of the anthracycline drug. This would account for the variation in the bound spermine molecule from one complex to the next. In spite of the differences in spermine conformation, several general characteristics of the interactions with DNA are constant throughout this series of complexes (figure 7). Each of the spermine molecules is located in the major groove and each spermine forms hydrophobic contacts with the DNA and the bound anthracycline. The positions of the DNA:anthracycline complexes which are involved in the hydrophobic interactions are highlighted in figure 8A. At least one methylene group per spermine molecule contacts a 5-methyl of thymine and a hydrophobic region of the anthracycline. These contacts exclude water molecules from contact with the hydrophobic regions of the DNA, the anthracycline and the spermine. From a comparison with DNA:anthracycline complexes without bound spermine molecules (44, 45), it appears that 5–7 water molecules are displaced from hydrophobic functionalities of the DNA upon binding of a single spermine molecule.

Z-DNA

The self-complementary DNA hexamer d(CGCGCG) crystallizes in the left-handed Z-conformation (46). The conformation of Z-DNA is distinct from other nucleic acid conformations and the major groove of B-DNA is replaced by a convex outer surface of Z-DNA. The minor groove of B-DNA is replaced by a deep groove of Z-DNA. The d(CGCGCG) hexamer was crystallized independently in the presence of magnesium and in the presence of spermine (37). Both the spermine and magnesium forms of these crystals diffract to atomic resolution (1.0 Å) allowing detailed examination of Z-DNA structure and the interactions of solvent, monocations and polyamines. The spermine form of Z-DNA contains two crystallographically distinct spermine molecules (figure 9).

One spermine molecule (spermine A) binds primarily to Z-DNA by forming a series of direct hydrogen bonds to the convex surfaces of two different (but symmetry related) Z-DNA duplexes. Each amino group of spermine A forms at least one hydrogen bond to the convex surface of a Z-DNA duplex. These hydrogen bonds are to the O6 and N7 positions of purines. These purines are located within three consecutive base pairs of one duplex and one base pair of another duplex. The first amino group of spermine A forms a hydrogen bond to the O6 of the guanine of the first base pair of one duplex. The third amino group of spermine A forms hydrogen bonds to both the O6 and N7 of the guanine of the third base pair of the same duplex. The fourth (terminal) amino group of spermine A interacts with the same duplex, forming a hydrogen bond to the N7 of the guanine of the second base pair and also to a phosphate oxygen on the backbone. The second amino group of spermine A forms a hydrogen bond to the N7 of a guanine of an adjacent duplex in the crystal.

The second spermine molecule (spermine B) interacts simultaneously with three different (but symmetry related) duplexes in the crystal. Spermine B forms only half

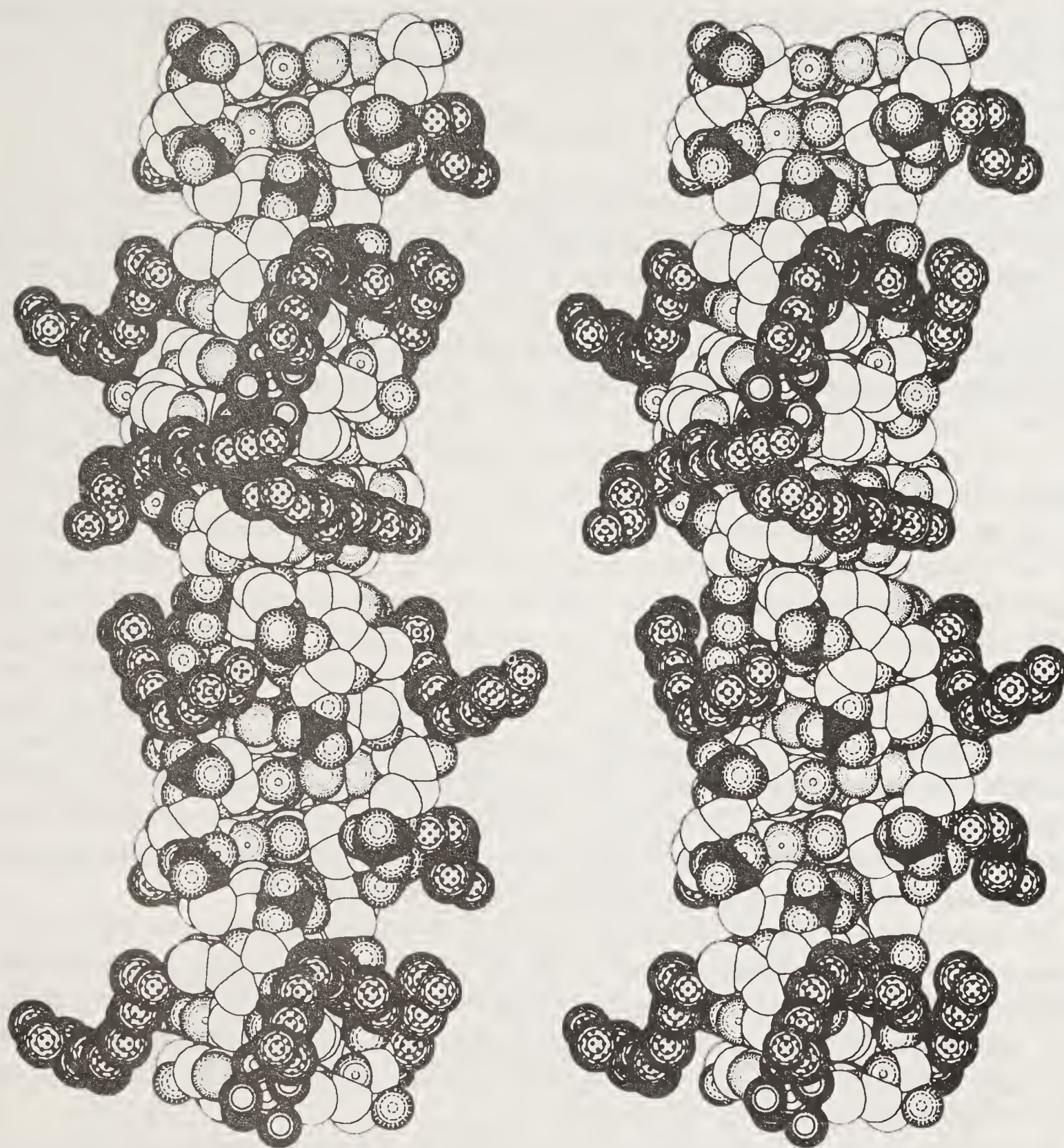


Figure 9. Stereoview van der Waals diagram of the spermine molecules bound to Z-DNA. The nitrogen atoms are dotted spheres, oxygen atoms are dashed spheres and phosphorus atoms as solid spheres. The spermine molecules are highlighted with dotted shading. [Reprinted from reference (37)].

as many hydrogen bonds to DNA bases as does spermine A but it forms twice as many hydrogen bonds to phosphate oxygens. Spermine B, like spermine A, forms hydrogen bonds to the O6 and N7 positions of purines and again these hydrogen bonds link the convex surfaces of two different duplexes in the crystal. The first two amino groups of spermine B form hydrogen bonds to guanine residues of adjacent base pairs on opposite strands of the same helix. Specifically, the terminal amino group of spermine B forms hydrogen bonds to both the N7 and the O6 of one guanine and the next amino group forms hydrogen bonds to the N7 of the adjacent guanine. Interacting with two additional duplexes, the other terminal amino group of spermine B forms hydrogen bonds to two phosphate oxygens of two additional duplexes in the crystal. Spermine B is located in a region where negatively-charged

phosphate oxygens form close intra- and intermolecular contacts binding in a mode reminiscent of the spermine found in yeast-tRNA^{Phe} crystals.

Discussion

The 3-dimensional X-ray crystallographic structure of spermine:nucleic acid complexes help to define the role of polyamines in nucleic acid structure and biology. Spermine binding does not appear to be exclusively delocalized-dynamic as predicted by the counterion condensation model (25) or static-specific as proposed by many modelling studies. Instead spermine binding combines aspects of both these modes of binding. The complexes described here, with tRNA (34), A-DNA (35), DNA:anthracycline complexes (32, 36) and Z-DNA (37), are clearly site-specific and static in the crystal as indicated by their X-ray crystallographic observation. The significance of site-specific interactions is further underscored by the series of DNA: anthracycline complexes where bound spermine is observed in some DNA:anthracycline complexes but not others, depending on DNA sequence. However, in solution one could imagine a spermine molecule flipping rapidly between the nearly isoenergetic bound states of the three spermine:DNA: anthracycline complexes (figure 8), without dissociating from DNA. Thus, in solution polyamine:nucleic acid complexes could be simultaneously site-specific and dynamic. Each crystallized spermine:nucleic acid complex appears to be stabilized predominantly by charge-charge interactions. Except for the consistency of this contribution, the mode of spermine binding to DNA is polymorphic. The hydrogen bonding and hydrophobic contributions to stability are highly variable and depend on the hydrophobic nature of the particular conformation or sequence of the DNA to which the spermine is bound.

The spermine:yeast-tRNA^{Phe} complex, with both spermine molecules located in regions between close contacts of negatively charged phosphate groups, first indicated crystallographically that charge-charge interactions are significant in spermine:nucleic acid complexes. Unfortunately, the importance of specific interactions such as hydrogen bonds and hydrophobic interactions cannot be definitively established at the resolution of the yeast-tRNA^{Phe} crystal.

In the spermine:A-DNA and spermine:DNA:anthracycline complexes, the spermine molecules bind in the grooves. The importance of charge-charge interactions in stabilizing these complexes is presumed from previous theoretical calculations indicating that floors of the grooves display the greatest electronegative potential of DNA (47-49). Favorable charge-charge interactions do not depend upon direct hydrogen bonds between positively charged amino groups and negatively charged phosphate oxygens. In X-ray crystal structures of DNA complexes, as described previously (32), positively charged amino groups of DNA binding ligands rarely form direct hydrogen bonds to phosphate oxygens of DNA. In the spermine:A-DNA and spermine:DNA:anthracycline complexes, methylene groups of spermine form hydrophobic contacts to 5-methyl groups of thymines and these complexes are stabilized significantly by hydrophobic interactions between the spermine and the DNA. The role of hydrogen bonding is variable with hydrogen bonds stabilizing the spermine:A-DNA complex to a greater extent than the spermine:DNA:anthracycline complexes.

In the spermine:Z-DNA complex, as in the yeast-tRNA^{Phe} complex, spermine molecules are located in regions of close contacts of negatively charged phosphate

groups and are similarly stabilized by charge–charge interactions. However, in contrast to the A-DNA and B-DNA anthracycline complexes, spermine molecules bound to Z-DNA do not form hydrophobic contacts with the DNA. Although alternating d(CG) lacks hydrophobic methyl groups, potential hydrophobic interaction is provided by the C-5 and C-6 positions of cytosines. Yet the methylene groups of the Z-DNA bound spermine molecules do not interact with cytosines but form occasional van der Waals contacts with hydrophilic sites of guanines. These contacts lack significant hydrophobic character. However, hydrophobic interactions are not unimportant in the spermine:Z-DNA complex as extensive contacts between the methylene groups of adjacent spermine molecules appear to provide a significant hydrophobic contribution to stability. In addition, the spermine molecules form many direct hydrogen bonds to DNA bases.

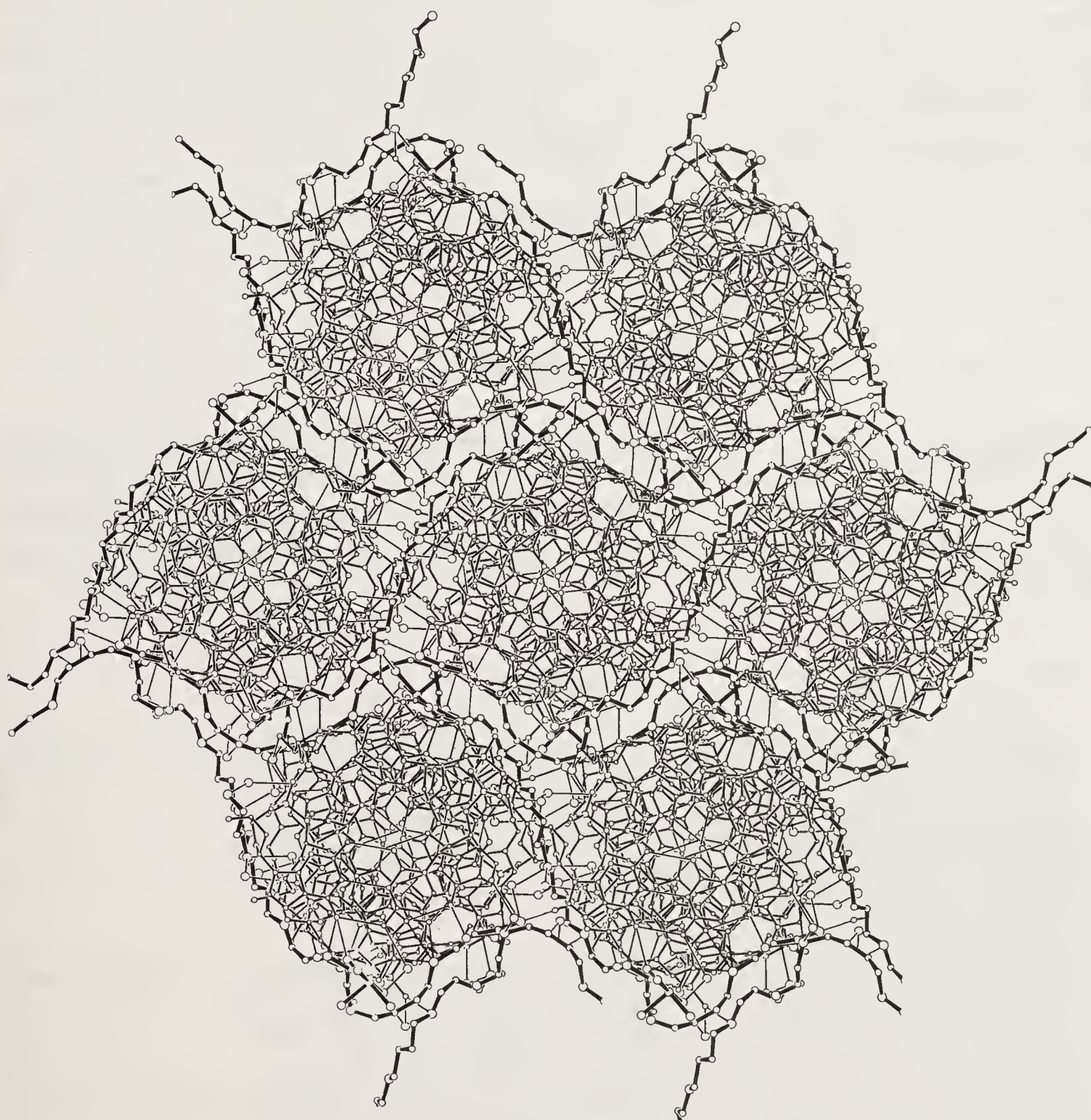


Figure 10. ORTEP diagram of Z-DNA:spermine complexes within the crystal, viewing down the helical axis. The DNA is drawn with thin lines and the spermine molecules are drawn with thick lines.

The role of polyamines in maintaining specific nucleic acid structural motifs can be best addressed by comparisons of nucleic acid X-ray structures that differ primarily by the presence or absence of bound polyamines. The best candidate for such a comparison is Z-DNA which has been crystallized independently in the presence of magnesium and in the presence of spermine (37). The spermine and magnesium forms of Z-DNA show only subtle differences in conformation. At least in the case of Z-DNA, magnesium can replace spermine without altering the conformation of the DNA. In a second example, spermine binds to some DNA:anthracycline complexes but not to others, depending on DNA sequence. In support of the Z-DNA results, the DNA conformation of the spermine form of the anthracycline complexes is very similar to the non-spermine form. This suggests that as the spermine fluctuates in solution, and even transiently dissociates from the DNA, the DNA remains relatively static.

The spermine molecules in the Z-DNA crystal provide a good model for polyamine condensation of DNA packaged within phage heads and in chromatin. In the Z-DNA crystal, the DNA approximates infinite parallel helices of DNA with negatively charged phosphates of adjacent helices in close proximity. Viewing down the helical axis (figure 10), spermine molecules form a cage-like arrangement around each Z-DNA helix, interacting simultaneously with several duplexes, effectively neutralizing negative charge, and stabilizing close contacts between duplexes.

The number of crystallized spermine:DNA complexes has increased recently and will no doubt continue to do so. New data will enhance our understanding of the structural roles carried out by these interesting molecules.

Acknowledgements

We thank Dr. Nassim Usman for helpful discussions. This research was supported by grants from the National Institutes of Health, the American Cancer Society, the National Science Foundation, the Office of Naval Research and the National Aeronautics and Space Administration. LDW acknowledges fellowship support from NIH.

References

1. Morris, D. R. (1981) in *Polyamines in Biology and Medicine*, eds. Morris, D. R. & Marton, L. J. (Marcel Dekker, New York) pp. 223–242.
2. Pegg, A. E. & McCann, P. P. (1982) *Am. J. Physiol.* **243**, C212–C221.
3. Tabor, C. W. & Tabor, H. (1984) *Annu. Rev. Biochem.* **53**, 749–790.
4. Schuber, F. (1989) *Biochem. J.* **260**, 1–10.
5. Mandel, M. (1962) *J. Mol. Biol.* **5**, 435–441.
6. Tabor, H. (1962) *Biochemistry* **1**, 496–501.
7. Bloomfield, V. A. & Wilson, R. W. (1981) in *Polyamines in Biology and Medicine*, eds. Morris, D. R. & Marton, L. J. (Marcel Dekker, New York) pp. 183–206.
8. Thomas, T. J. & Bloomfield, V. A. (1984) *Biopolymers* **23**, 1295–1306.
9. Gosule, L. C. & Schellman, J. A. (1978) *J. Mol. Biol.* **121**, 311–326.
10. Chatteraj, D. K., Gosule, L. C. & Schellman, J. A. (1978) *J. Mol. Biol.* **121**, 327–337.
11. Wilson, R. W. & Bloomfield, V. A. (1979) *Biochemistry* **18**, 2192–2196.
12. Widom, J. & Baldwin, R. L. (1980) *J. Mol. Biol.* **144**, 431–453.
13. Crothers, D. & Sen, D. (1986) *Biochemistry* **25**, 1495–1503.
14. Sen, D. & Crothers, D. M. (1986) *Biochemistry* **25**, 1495–1503.

15. Behe, M. J. & Felsenfeld, G. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1619–1623.
16. Chen, H. H., Behe, M. J. & Rau, D. C. (1984) *Nucleic Acids Res.* **12**, 2381–2389.
17. Rao, M. V. R., Atreyi, M. & Saxena, S. (1990) *Biopolymers* **29**, 1495–1497.
18. Geiger, L. E. & Morris, D. R. (1980) *J. Bacteriol.* **141**, 1192–1198.
19. Pingoud, A. (1985) *Eur. J. Biochem.* **147**, 105–109.
20. D'Orazi, D., Fracassini, D. S. & Bagni, N. (1979) *Biochem. Biophys. Res. Commun.* **90**, 362–367.
21. Morgan, J. E., Calkins, C. C. & Matthews, H. R. (1989) *Biochemistry* **28**, 5095–5106.
22. Garcia, A., Giege, R. & Behr, J.-P. (1990) *Nucleic Acids Res.* **18**, 89–95.
23. Braunlin, W. H., Strick, T. J. & Record, M. T., Jr. (1982) *Biopolymers* **21**, 1301–1314.
24. Wemmer, D. E., Srivenugopal, K. S., Reid, B. R. & Morris, D. R. (1985) *J. Mol. Biol.* **185**, 457–459.
25. Manning, G. S. (1978) *Q. Rev. Biophys.* **11**, 179–246.
26. Tsuboi, M. (1964) *Bull. Chem. Soc. Jpn.* **37**, 1514–1522.
27. Liquori, A. M., Costantino, L., Crescenzi, V., Elia, V., Giglio, E., Puliti, R., de Santis Savino, M. & Vitagliano, V. (1967) *J. Mol. Biol.* **24**, 113–122.
28. Burton, D. R., Forsen, S. & Reimarsson, P. (1981) *Nucleic Acids Res.* **9**, 1219–1228.
29. Feuerstein, B. G., Pattabiraman, N. & Marton, L. J. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5948–5952.
30. Zakrzewska, K. & Pullman, B. (1986) *Biopolymers* **25**, 375–392.
31. Feuerstein, B. G., Basu, H. S. & Marton, L. J. (1988) in *Progress in Polyamine Research*, eds. Zappia, V. & Pegg, A. E. (Plenum Press, New York) pp. 517–523.
32. Williams, L. D., Frederick, C. A., Ughetto, G. & Rich, A. (1990) *Nucleic Acids Res.* **18**, 5533–5541.
33. Iitaka, Y. & Huse, Y. (1965) *Acta Crystallogr.* **18**, 110–121.
34. Quigley, G. J., Teeter, M. M. & Rich, A. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 64–68.
35. Jain, S., Zon, G. & Sundaralingam, M. (1989) *Biochemistry* **28**, 2360–2364.
36. Frederick, C. A., Williams, L. D., Ughetto, G., van der Marel, G. A., van Boom, J. H., Rich, A. & Wang, A. H.-J. (1990) *Biochemistry* **29**, 2538–2549.
37. Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A. & Wang, A. H.-J. (1989) *J. Biol. Chem.* **264**, 7921–7935.
38. Drew, H. R. & Dickerson, R. E. (1981) *J. Mol. Biol.* **151**, 535–556.
39. Kim, S. H., Quigley, G. J., Suddath, F. L. & Rich, A. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 841–845.
40. Ichikawa, T. & Sundaralingam, M. (1972) *Nature, New Biology* **236**, 174–175.
41. Rich, A. & RajBhandary, U. L. (1976) *Annu. Rev. Biochem.* **45**, 805–860.
42. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, (Springer-Verlag, New York).
43. Acramone, F. & Penco, S. (1989) in *Antitumor Natural Products*, eds. Takeuchi, T., Nitta, K. & Tanaka, N. (Japan Scientific Press, Tokyo) pp. 1–53.
44. Williams, L. D., Egli, M., Ughetto, G., van der Marel, G. A., van Boom, J. H., Quigley, G. J., Wang, A. H.-J., Rich, A. & Frederick, C. A. (1990) *J. Mol. Biol.* **215**, 313–320.
45. Wang, A. H.-J., Ughetto, G., Quigley, G. J. & Rich, A. (1987) *Biochemistry* **26**, 1162–1163.
46. Rich, A., Nordheim, A., & Wang, A. H.-J. (1984) *Annu. Rev. Biochem.* **53**, 791–846.
47. Lavery, R. & Pullman, B. (1981) *Nucleic Acids Res.* **9**, 4677–4688.
48. Weiner, P. K., Langridge, R., Blaney, J. M., Schaefer, R. & Kollman, P. A. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3754–3758.
49. Lavery, R. & Pullman, B. (1985) *J. Biomol. Struct. Dynam.* **2**, 1021–1032.

Stereochemical Studies on DNA and DNA–Drug Interactions[†]

S. Mohan and N. Yathindra*

Department of Crystallography and Biophysics, University of Madras,
Guindy Campus, Madras 600 025, India

Introduction

The resurgence of activities in the last decade on DNA structural research has provided a wealth of evidence towards the nonuniform nature of DNA structure, with a tendency to exhibit sequence dependent conformational heterogeneity, and the ability of DNA to assume unusual structures. Perhaps one of the most important and difficult questions yet unanswered is the extent to which these can be attributed to intrinsic factors like base sequence, external factors like binding of drugs, proteins etc., and how they can be correlated as sequence specific molecular recognition features. While there exists some evidence concerning the influence of both intrinsic and extrinsic factors, precise correlation between base sequence and DNA structure must await more detailed results from a large variety of sequences, especially in view of lack of understanding of the role of adjacent as well as near-neighbour bases.

Stereochemical approaches to probe biopolymer structure and conformation through computer modelling have proved important and often stimulated experimental research. We have been using these techniques recently to examine certain sequence dependent fine structural properties in DNA helices, binding of drugs to DNA and effect of modification of sugar phosphate backbone on DNA conformations and their transitions. Here, we present results on some novel cross strand hydrogen bonds possible between Watson-Crick paired bases of DNA helices and interactions of a groove binding molecule with DNA.

Method

Mathematical procedures based on Eulerian motions are used to generate successive base pairs and also to effect variation in their relative orientations and positions. Same procedures are also adopted in molecular docking involved in interaction studies. Conformational energy calculations are performed using the molecular mechanics program AMBER.

[†]Contribution Number 772 from this department

*For correspondence

Results and Discussion

Propeller Twist Driven Sequence Dependent Cross Strand Hydrogen Bonds

One of the most common and conspicuous structural features observed in both A- and B-DNA helices is the propeller twist in the Watson–Crick base pairs (1). While propeller twists of varying magnitudes are observed for A–T as well as G–C pairs, the nature of influence of adjacent and near-neighbour bases on their magnitudes are not clear. An interesting consequence of the propeller twist has been the possibility of a novel bifurcated hydrogen bond between cross strand bases as first observed in the crystal structure containing a stretch of A bases (2). It has been argued (2) that this extra hydrogen bond stiffens the polynucleotide chain and therefore may offer explanations for the inability of poly(dA).poly(dT) to wrap around the nucleosomal core and also for the observed curvature in A-tract containing DNA sequences (3). The fact that non-homopolymeric base pair juxtapositions also possess similar hydrogen bond donor and acceptor sites prompted us (4, 5) to search for other likely sequence combinations that may promote such cross strand and possibly other types of hydrogen bonds in DNA duplexes. Both computer aided simulations of models and conformational energy calculations (6) are carried out on all the base pair doublets, initially with base pairs alone and subsequently extended to duplex dimers and trimers which include sugar–phosphate backbone.

Models Involving Base Pairs

Among the 16 base pair doublet sequences shown in table 1, only ten turn out to be unique due to symmetry and these alone are considered in the investigation. The starting helical geometries of all the base pair doublets correspond to B-DNA (7). Search for cross strand hydrogen bond (CSH) was carried out by varying the fine structural parameters such as propeller twist, roll, tilt, buckle, slide etc., in addition to helical parameters according to Cambridge convention (8). The parameters were optimised for the best bifurcated hydrogen bond geometry. The distance of $< 2.4 \text{ \AA}$ for H...O and a hydrogen bond angle of $\geq 90^\circ$ together with less than 0.3 \AA deviation of the hydrogen atom from the plane formed by the two acceptors and the donor are considered as optimal for a three-centered hydrogen bond (9, 10). It is suggested (10) that a longer H...O distance of upto 3.0 \AA is effective in view of long range influence of electrostatic interactions. It is found from the analysis that out of the ten combinations of different base pair juxtapositions only five exhibit propensity for CSH bonds. Occurrence of like group oppositions in the other three combinations does not permit CSH but on the other hand results in steric clash consequent to propeller twists. The sequence combinations promoting CSH are shown prominently by enclosing them by thick lines in table 1. Possibility of CSH in both major and minor groove exists and is indicated by the labels as M and m along the arrows which point from the donor to the acceptor base.

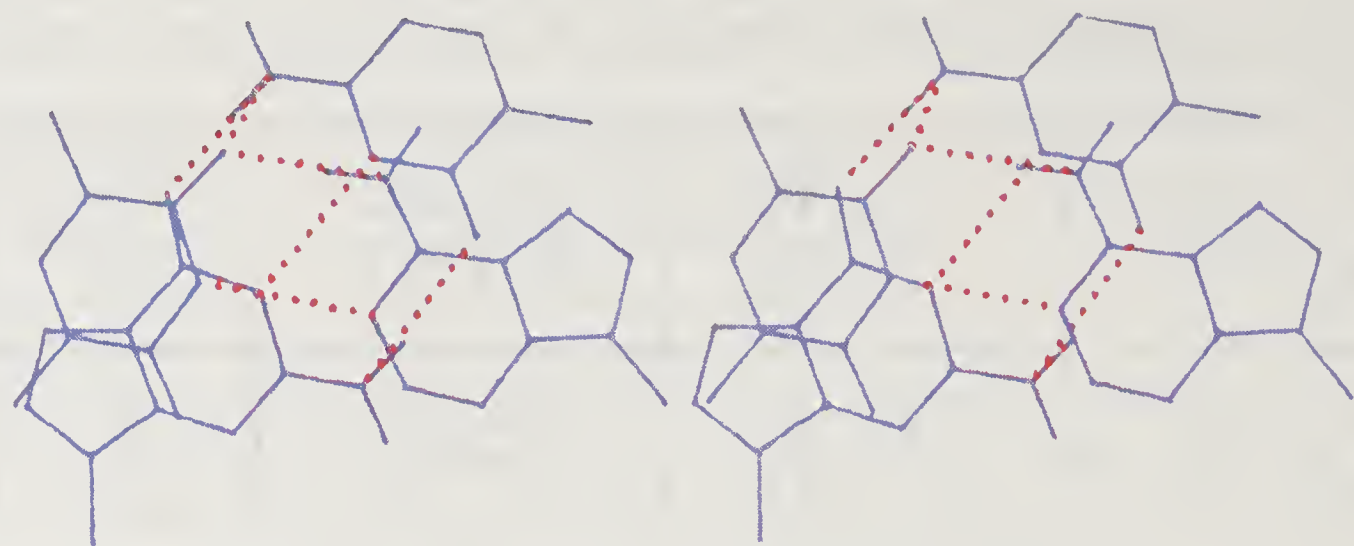
Stereo views of various possible bifurcated cross strand hydrogen bonds in different base pair doublets are shown in figure 1. The different base pair parameters that are required to facilitate their formation are given in table 2. Besides the prerequisite propeller twists, the occurrence of CSH necessitates varying amounts of roll angle (-5° to 11°), slide (-0.5 to 1.6 \AA), buckle (-6° to 14°) etc., depending upon the combinations of base pair sequences. The helical parameters also vary in

Table 1. The 16 possible dimer sequences. The sequences that promote CSH are enclosed by thick lines. The CSH in major and minor grooves are represented by 'M' and 'm' on the arrows which point from donor to acceptor. Bases at the left hand bottom corner in every doublet correspond to the 5'-end.

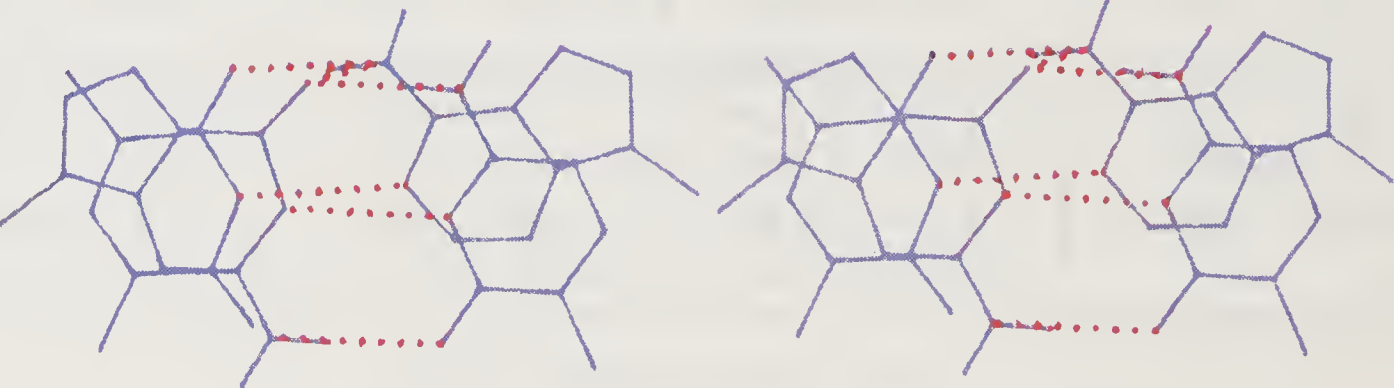
the range, 28–40° for the helical twist, and 3.0–3.4 Å for the helical rise. It is most interesting that adjacent CG base pairs with fairly low propeller twist can also exhibit a CSH in the major groove similar to that found in the homopolymeric A-stretch. In addition, it is also possible to visualise a CSH in the minor groove with a small negative roll angle. It may be mentioned that these hydrogen bonds appear quite naturally in a B-type helix. The characteristic large slide (negative Y-slide) between adjacent base pairs in A-type helix renders the potential CSH bonding groups to be considerably displaced and therefore they are less likely in canonical A-DNA helices (7). Therefore it is not surprising that such a CSH is not observed in any of the several A-DNA crystal structures comprising adjacent GC base pairs (11,12). For similar reasons CSH in adjacent AT base pairs is likely to be favorable only in B-DNA helices.

On the other hand a slide becomes necessary for the formation of major groove CSH in certain sequences like GT/AC and TG/CA. The occurrence of 'large' adenine base necessitates only a small slide in the same direction as in A-type helix and this helps in bringing to proximity the hydrogen bonding groups in GT/AC resulting in a CSH between G and A. While the requirement of a slide and a low helical twist ($< 30^\circ$) seem to suggest that this CSH may be favored in A-type helices, calculations suggest that low values of slide (-0.5 to -0.75 Å) are sufficient to facilitate the formation of this purine...purine CSH. Since this range of slide is also observed in B-type helices (12), it is likely that this hydrogen bond may be feasible in both A and B type helices as well. However, it is found that helical twist needs to be low since values ($\geq 36^\circ$) render the hydrogen bonding groups to be in positions not conducive for optimal bifurcated hydrogen bond formation. A low buckle angle with a higher slide and a propeller twist, and a high buckle angle concomitant with

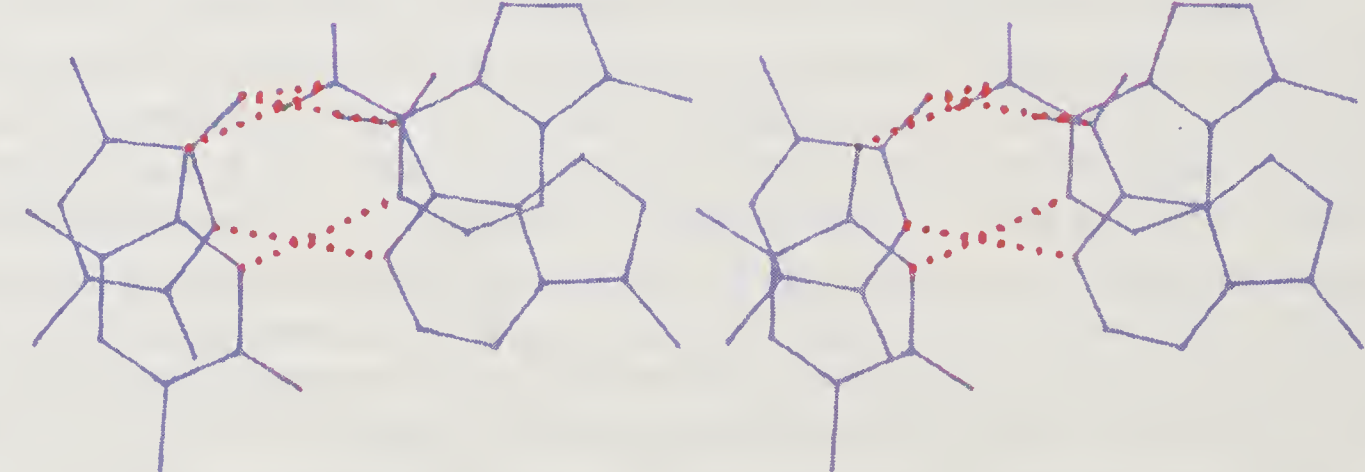
(a)



(b)



(c)



(d)

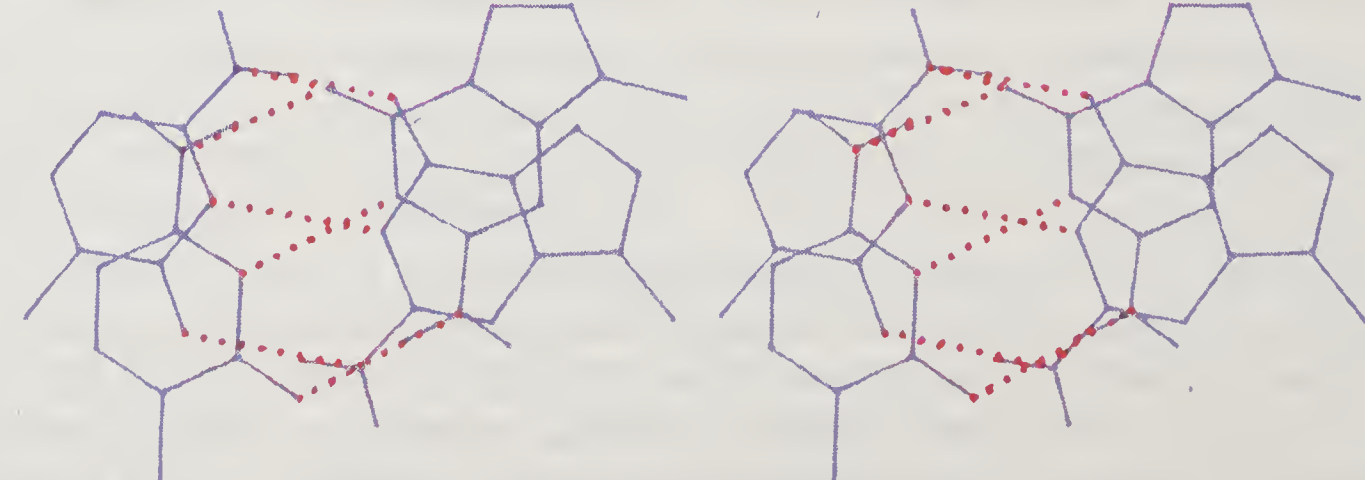


Figure 1. *Contd.*

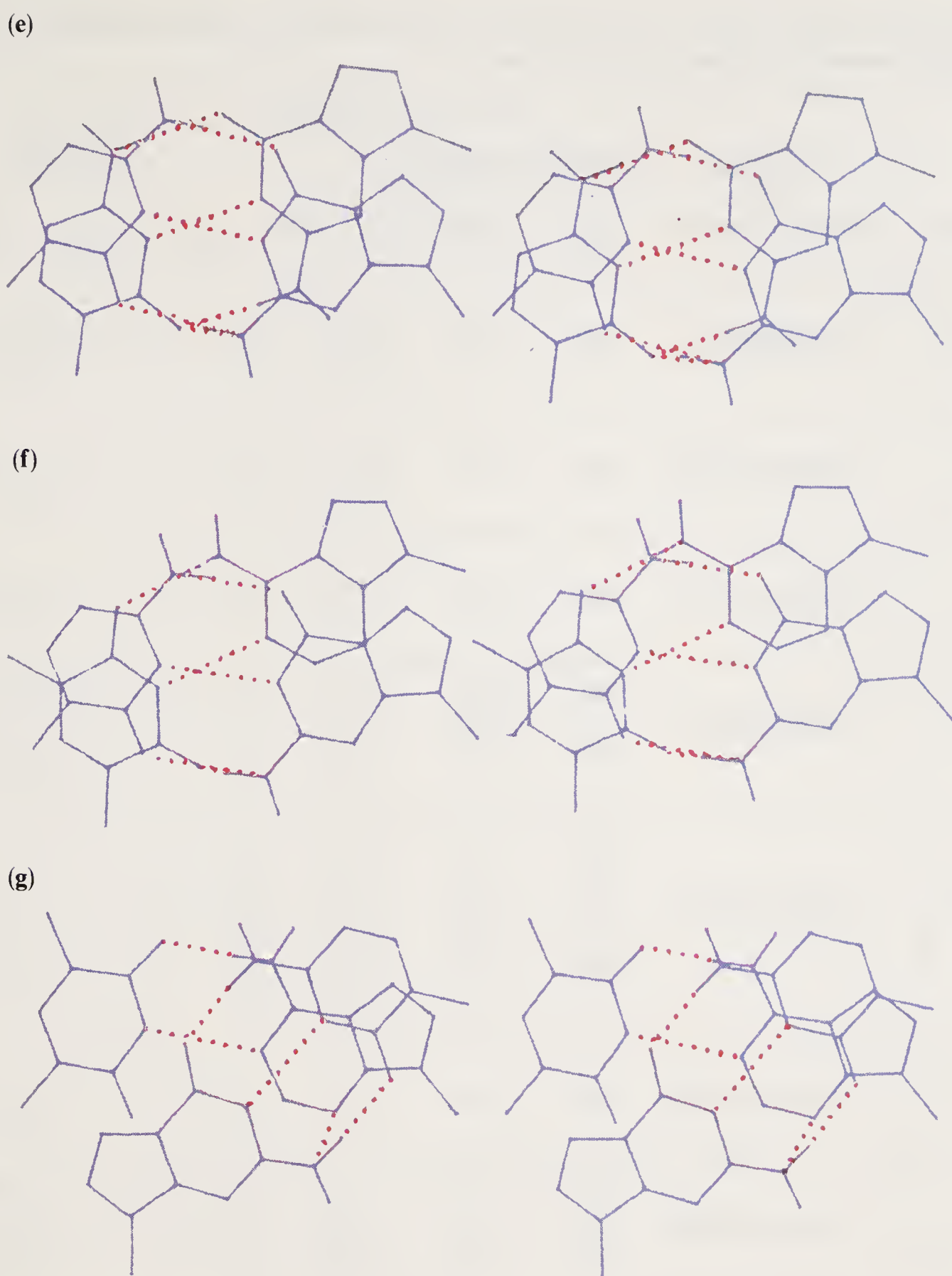


Figure 1. Stereo views of different base pair dimer sequences exhibiting CSH between them. Hydrogen bonds on the major groove side (a–d) and minor groove side (e–g) are shown by dotted lines. (a) CSH between (C) N4 & (T)O4 in TG/CA in the major groove. (b) CSH between (G)O6 & (A)N6 in GT/AC in the major groove. (c) CSH between (T)O4 & (A)N6 in TT/AA in the major groove. (d) CSH between (C)N4 & (G)O6 in CC/GG in the major groove. (e) CSH between (C)O2 & (G)N2 in CC/GG in the minor groove (f) CSH between (T)O2 & (G)N2 in CT/AG in the minor groove. (g) CSH between (G)N2 & (A)N3 in TG/CA in the minor groove.

a low slide and a propeller twist can result in this CSH with essentially similar hydrogen bond geometry.

On the other hand, if the above sequence is reversed as in TG/CA, the G...A CSH observed above is not found to be feasible since a negative propeller twist renders the hydrogen bonding groups to be distal. However, this brings forth the possibility

Table 2. Helical and basepair parameters that promote CSH in various basepair doublets. Cambridge convention representations are used.

Sequence	Hydrogen bond Donor/Acceptor	Helical Parameters			Base Pair Parameters		
		$\Omega(^{\circ})$	Dz(Å)	$\omega(^{\circ})$	$\rho(^{\circ})$	$\tau(^{\circ})$	$\kappa(^{\circ})$
<div>C → G C ← G</div>	N4(C1)/O6(G3)	36	3.0	CG1 = − 18 CG2 = − 14	10	1.0	CG1 = − 2 CG2 = − 2
<div>C → G C ← G</div>	N2(G4)/O2(C2)	34	3.4	CG1 = − 14 CG2 = − 14	− 6		CG1 = − 4 CG2 = − 2
<div>T → A C ← G</div>	N2(G4)/O2(T2)	31	3.2	CG = − 14 TA = − 14	− 5		CG = 0 TA = 0
<div>T → A G ← C</div>	N6(A3)/O6(G1)	28 30	3.2 3.0	GC = − 20 TA = − 20 GC = − 8 TA = − 16	11 6		GC = 4 TA = 4 GC = 14 TA = 8
<div>G → C T ← A</div>	N4(C3)/O4(T1)	28 34 40	3.0 3.2 3.4	TA = − 18 GC = − 18 TA = − 16 GC = − 16 TA = − 20 GC = − 20	9 10 10		TA = 2 GC = − 2 TA = 0 GC = − 6 TA = 4 GC = 4
<div>T → A T ← A</div>	N6(A3)/O4(T1)	35	3.1	TA1 = − 22 TA2 = − 22	7		TA1 = 4 TA2 = 4
<div>G → C T ← A</div>	N2(G2)/N3(A4)	30 40	3.4 3.4	GC = − 18 TA = − 18 GC = − 18 TA = − 18	− 8 − 5		GC = − 2 TA = 2 GC = − 2 TA = 2

Ω = Helical Twist: Dz = Rise/base pair: ω = Propeller Twist: ρ = Roll: τ = Tilt: κ = Buckle: D_x = X-slide: D_y = Y-slide: D = Deviation of H from the plane formed by two acceptors and one donor

CSH-Bond Geometry					
$D_x(\text{\AA})$	$D_y(\text{\AA})$	Angle ($^\circ$)	Length(\AA)		$D(\text{\AA})$
			N \cdots N/O	O \cdots H	
-0.3	0.2	103.53	2.84	2.43	0.07
0.5	0.5	103.47	2.71	2.29	0.04
0.5	0.5	106.08	2.74	2.29	0.04
0.2	-0.75	101.51	2.90	2.54	0.09
.....
0.5	-0.5	101.19	2.86	2.50	0.13
1.1	1.8	104.4	2.75	2.36	0.02
.....
0.3	1.1	96.80	2.76	2.47	0.01
.....
0.5	1.5	94.00	2.76	2.50	0.15
0.5	0.2	103.1	2.79	2.38	0.10
-1.3	0.0	101.2	2.86	2.45	0.16
.....
-1.3	0.0	98.3	2.73	2.47	0.12

of a new CSH, also in the major groove, but involving now the cross strand pyrimidines. This hydrogen bond between C and T is characterised by a positive Y-slide (1.1 to 1.8 Å) i.e., opposite to that found in A-type helices and that required for the purine...purine (G-A) CSH discussed above. Slide in X-direction is also found to be necessary. Magnitude of the slide required seems to be determined by the helical twist (28–40°), large X-slide (shift) being associated with low helical twist for the occurrence of CSH. Formation of both these hydrogen bonds simultaneously seems natural in a trimer sequence comprising a T-A pair flanked by G-C pair on either side since a slide for the central T-A satisfies both the requirements. While this pyrimidine...pyrimidine bifurcated hydrogen bond seems difficult for A-type helices with negative propeller twists especially in view of positive slide, the nature of the slide and the helical twist required are reminiscent of features seen in C-DNA like helix (12). Therefore these CSH bonds are most likely to co-exist in B-type rather than A-type helices. A minor groove CSH is also possible for the sequence CT/AG involving N2 of G and O2 of T and the base pair parameters for optimal hydrogen bonding are very similar to those of the minor groove CSH seen between adjacent G-C pairs. Another minor groove CSH for the sequence TG/CA is realised with a negative X-slide and a negative roll angle (–8 to –10°). This involves N3 of adenine and N2 of guanine and is of the N–H...N type. Thus, there are about 7 possibilities of CSH including that observed first in adjacent A-T pairs (2). Some of these involve purine...purine, purine...pyrimidine and pyrimidine...pyrimidine and occur in the major and minor groove sides of helices. The results of the analysis also show that a majority of these CSH hydrogen bonds favor B-type helical geometry for the adjacent base pairs with varying amounts of negative propeller twists. Apparently, the slide between adjacent base pairs inherent to canonical A-type helices (7) render the CSH possibility unlikely in A-type helices with the exception of purine...purine

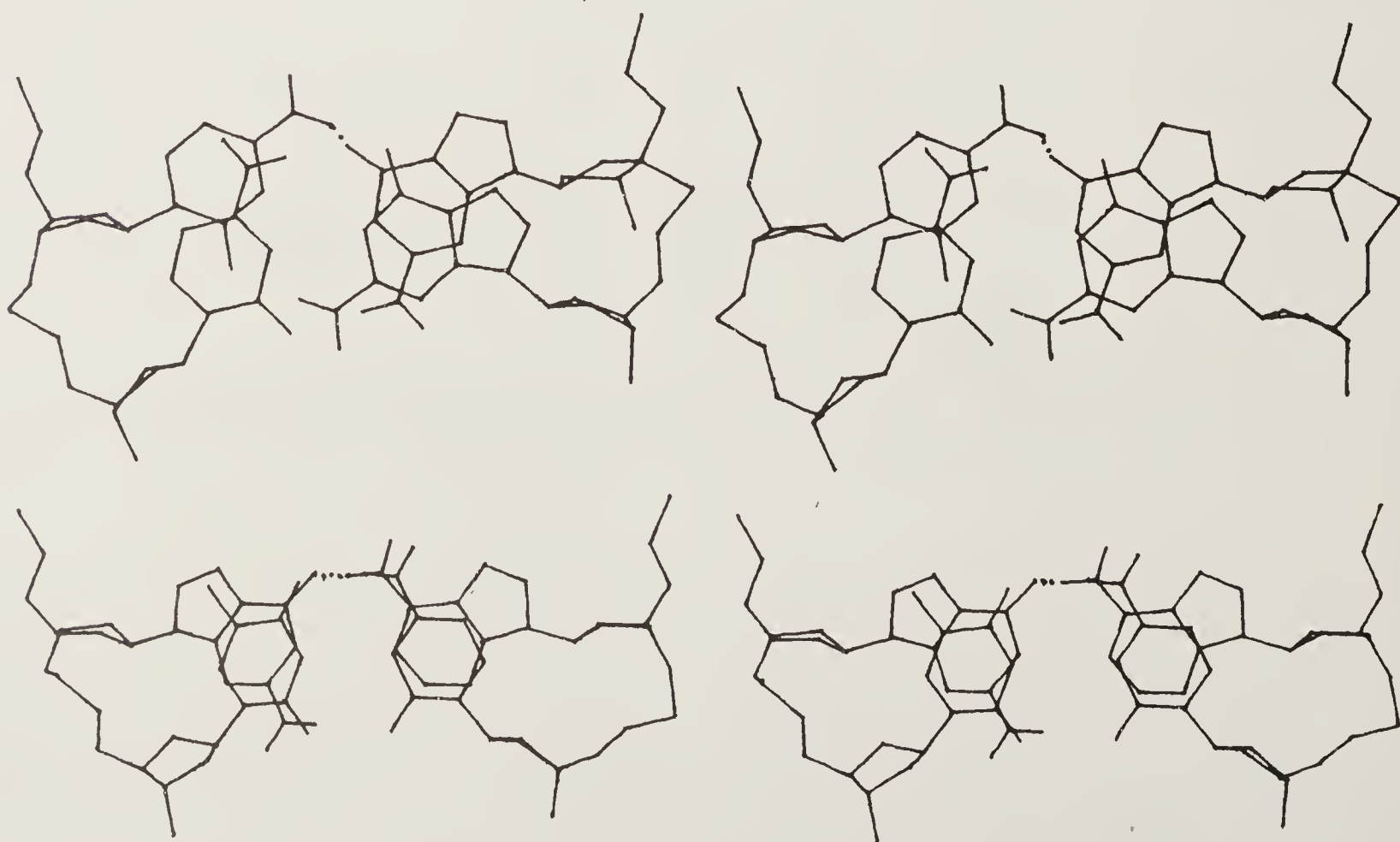


Figure 2. Stereo views of two representative dimer sequences CC/GG and TG/CA exhibiting CSH. Backbones fixed (B-DNA geometry) between these base pairs without disturbing the base pair parameters (27).

(G...A) CSH, where the large size of the bases seemingly compensate for the effect of slide. Thus, the results clearly show that CSH are not only sequence dependent but also depend on whether the helix is of A or B-DNA type.













Although the above analyses are performed considering only the base pairs, their relevance to nucleic acids has been demonstrated by retaining the base-pair parameters and hence the CSH even after affixing sugar–phosphate backbones. Figure 2 shows such a representative duplex dimer exhibiting CSH. While the parameters shown in table 1 are only representative, it is possible that in reality interactions involving sugar–phosphate backbone and other external factors may modify some of them.

Molecular Mechanics Studies on Duplex Dimers and Trimers

In order to verify if such CSHs are also energetically favored notwithstanding the conformational constraints of helical sugar–phosphate backbone, molecular mechanics calculations are carried out using AMBER on all the above five duplexes comprising different base pair sequences. The results are shown in table 3. The initial geometries of all the five dimer duplexes correspond to fiber B-DNA (7) wherein every base pair is propeller twisted by nearly $+4^\circ$. It is readily observed that all the base pairs after minimization take up negative propeller twists in the range -1.0° to -16.3° . The magnitude of the propeller twist seems to be dictated by adjacent base sequences and this has a bearing on the occurrence as well as strength of CSH. Stereoviews of energy minimised duplex dimers exhibiting CSHs are shown in figure 3.

It is clear from the hydrogen bond parameters (table 3) that adjacent C–G pairs alone exhibit a strong CSH suggesting that homopolymeric C–G stretch also possesses this property similar to homopolymeric A–stretch and may contribute in further stiffening of the already stable poly (dG).poly(dC) helix. It is therefore also important

Table 3. Propeller twists and CSH geometry in energy minimized duplex dimers. Arrows have the same meaning as in table 1.

Dimer Sequence	Propeller Twist ($^\circ$)	Bifurcated C.S.H-Bond Geometry		
		O...H (Å)	N...O (Å)	Angle ($^\circ$)
T  A T  A	– 11.61 – 16.35	2.98	3.30	101.5
C  G C  G	– 9.68 – 16.23	2.35	3.07	127.8
T  A G  C	– 4.09 – 1.00	2.71	2.94	92.6
G  C T  A	– 11.55 – 11.60	3.07	3.70	123.3
T  A C  G	– 8.70 – 13.85	2.99	3.08	83.0
C  G C  G	– 9.68 – 16.23	2.98	3.2	87.1

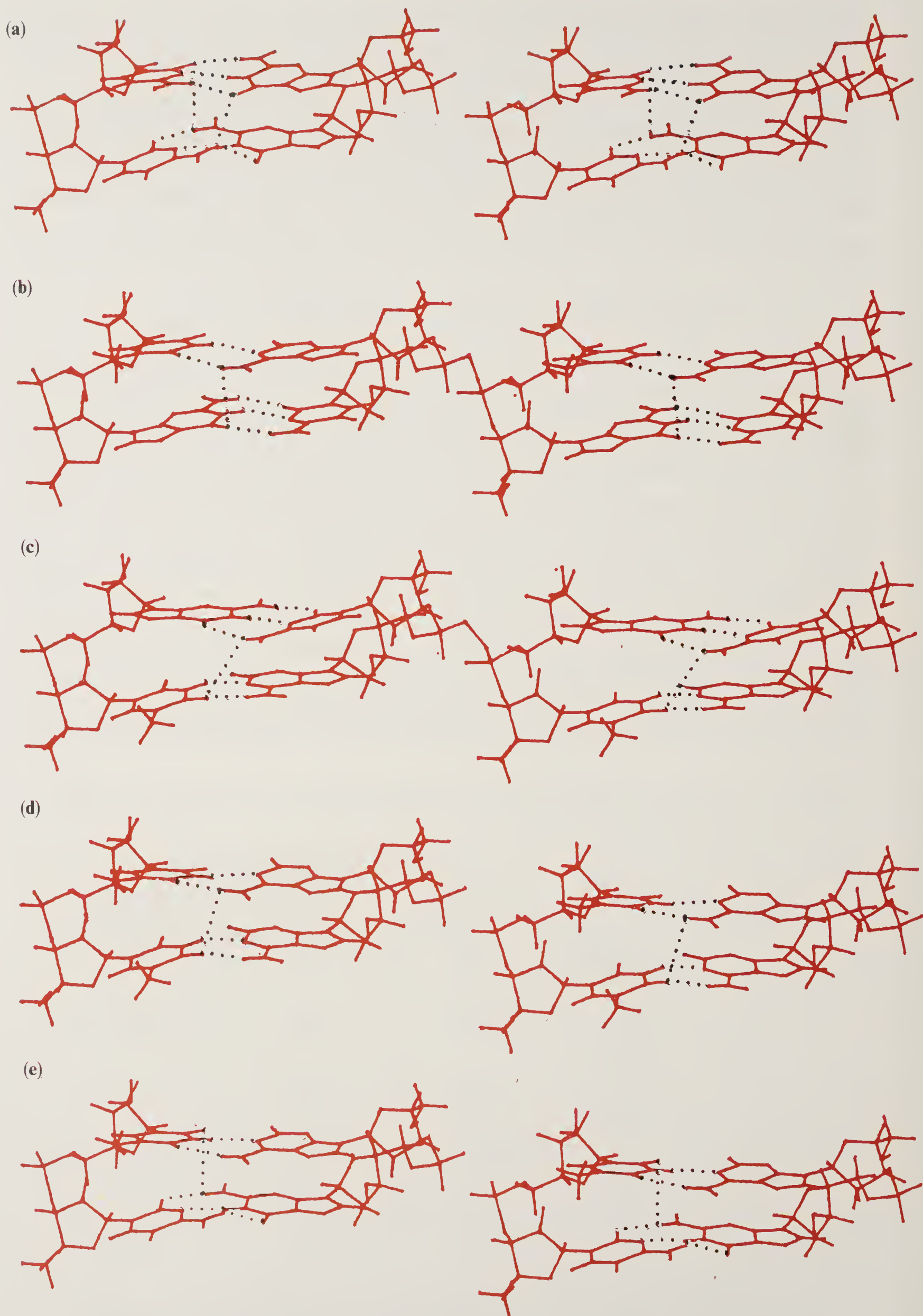


Figure 3. Stereo plots of energy minimized duplex dimers showing CSH. (a) CC/GG (b) GT/AC (c) TG/CA (d) TT/AA (e) CT/AG.

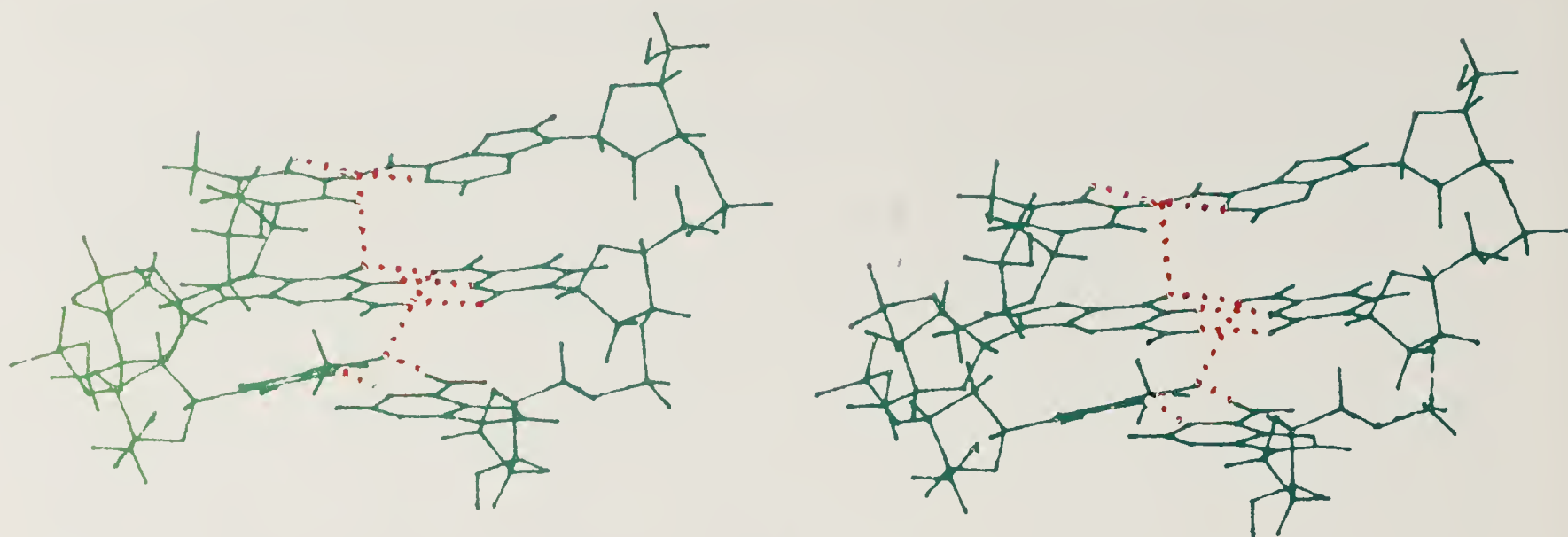
to probe the influence of such G-tract on the bending properties of DNA and compare them with those found with A-tract (3).

In all the other sequences, H...O distances defining the CSH vary from 2.7 to 3.0 Å while satisfying other criteria. The reason for the observed long O...H distance for most of the CSHs deduced from molecular mechanics calculations is that other base pair parameters like slide, buckle, tilt etc., apart from the propeller twist are nearly zero. Even minor variation in these will lead to better geometries for the hydrogen bond as clearly shown from model studies discussed above. It has been argued (10) recently that H...O distances upto 3.0 Å can be regarded as an effective hydrogen bond because of the long range nature of coulombic interactions. Accordingly therefore all the major groove CSHs predicted to occur in different sequences by model studies (4, 5) are in fact shown to be present in the energy minimized duplexes although the CSH are not as strong. It may be noted that the homopolymeric CC/GG sequence exhibits a weak CSH on the minor groove side with a H...O distance of 2.98 Å, in addition to the strong major groove hydrogen bond. Similarly, of the two possible minor groove cross strand hydrogen bonds, one between G and T in the sequence CT/AG suggested from considerations of base pairs alone is noticed whereas the other involving N3 of adenine and N2 of guanine in the sequence CA/TG is not observed.

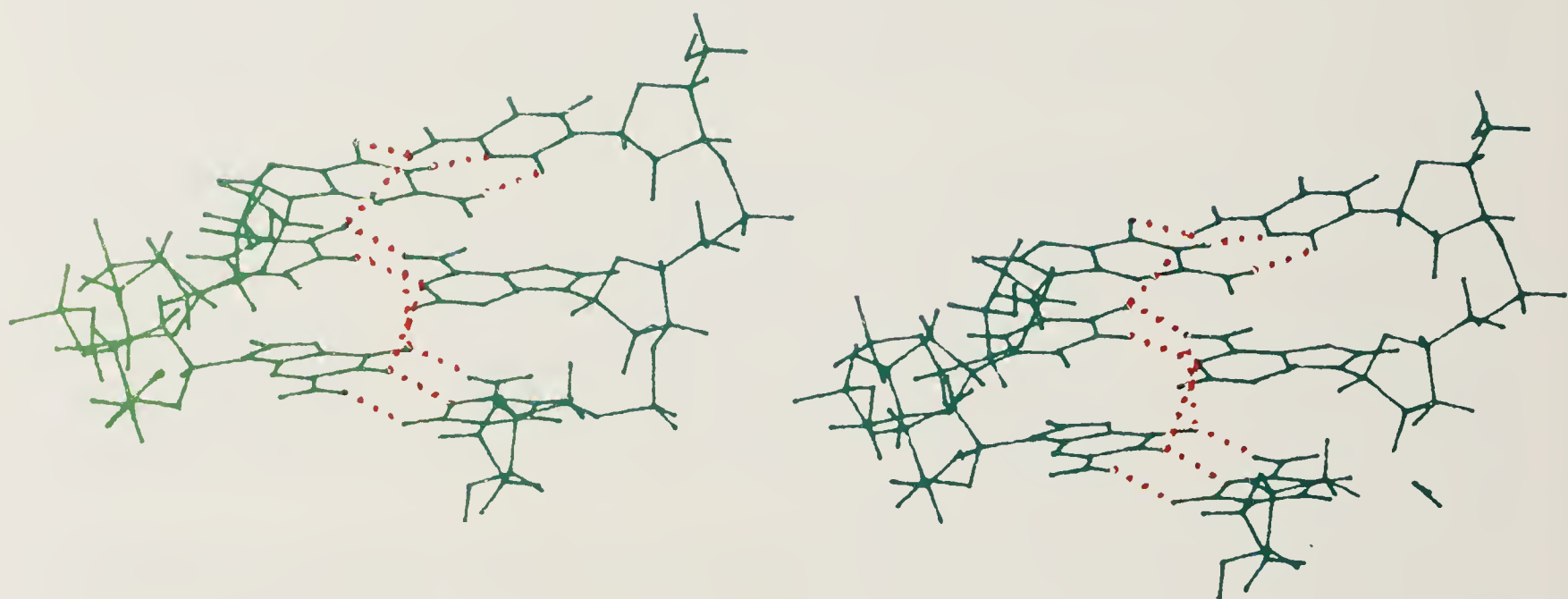
Another interesting possibility of bifurcated hydrogen bond that emerges from molecular mechanics calculations on dimer duplexes is that involving guanine N1 and adenine N1 in the sequence GT/AC and thymine N3 and adenine N1 in AA/TT. These CSHs are of the N–H...N type and the H...N distances range from 2.77 to 3.06 Å and occur at the centre of helix. There is yet another type of bifurcated hydrogen bond involving guanine N1 and cytosine O2. This is not a cross strand hydrogen bond but occurs within a GC base pair and is found in all the GC base pairs irrespective of their juxtaposition.

Calculations have also been extended to duplexes with three base pairs to examine the co-existence of different CSHs as in a DNA fragment since model studies implied different base pair parameters for the occurrence of different CSHs. Besides, base paired trinucleoside phosphates represent the shortest stretch needed to investigate the mutual influence of adjacent base pairs on their properties including propeller twist. All the 12 unique trimer sequences (out of the 64 possible triplet combinations) derived from various combinations of duplexes showing propensity for CSH at consecutive steps were subjected to molecular mechanics study. The results are shown in table 4. As expected all the trimer sequences demonstrate a clear tendency to form CSH at every step suggesting that CSHs can in fact co-exist along a DNA chain containing appropriate sequences. They are not as strong except in the homopolymeric G stretch, for reasons already mentioned. One interesting observation is that some trimer sequences such as CCT/AGG, CTT/AAG, CTG/CAG and ACT/AGT are involved in one major and one minor groove CSH. Stereoviews of duplex trimers exhibiting such CSHs are shown in figure 4. In such cases the same base of the central base pair is involved in the formation of both the CSHs.

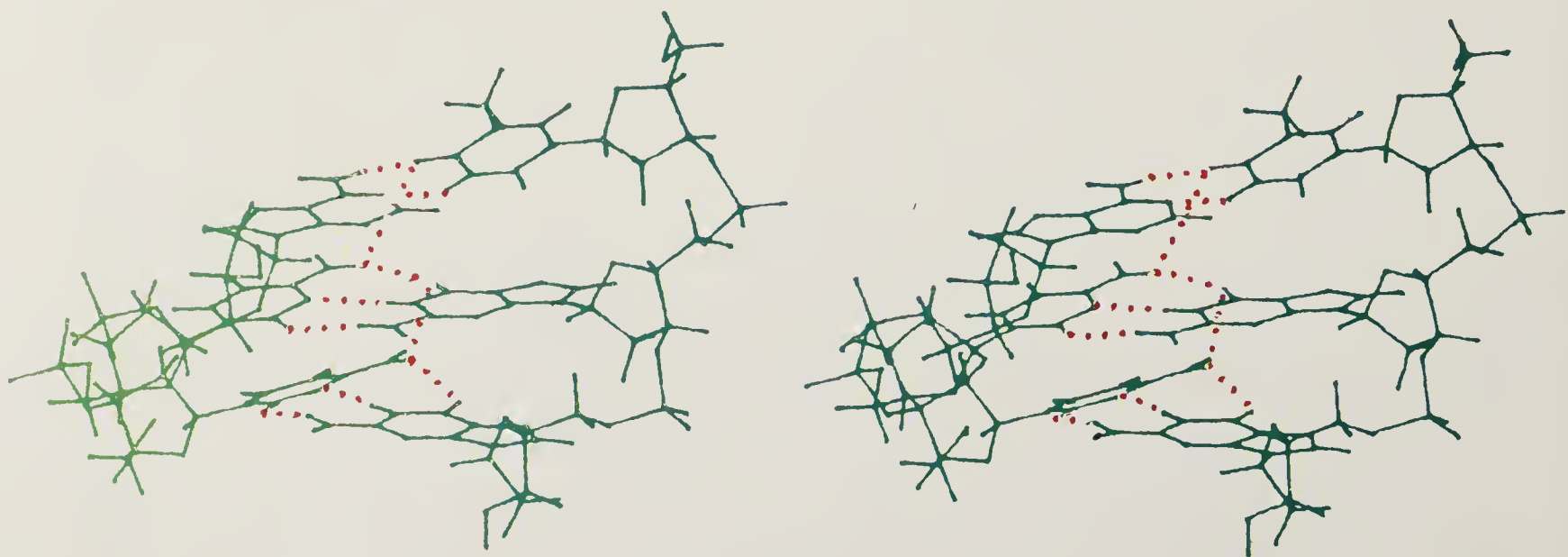
It is found in general that the CSHs on the major groove side take up better geometry. This could also imply that a negative roll angle which is a prerequisite for minor groove CSH is probably less favored compared to positive roll angle. The striking result that emerges out of these studies is that not only the negative propeller twist of base pairs but also the CSHs between propeller twisted base pairs seem to



(a)



(b)



(c)

Figure 4. *Contd.*

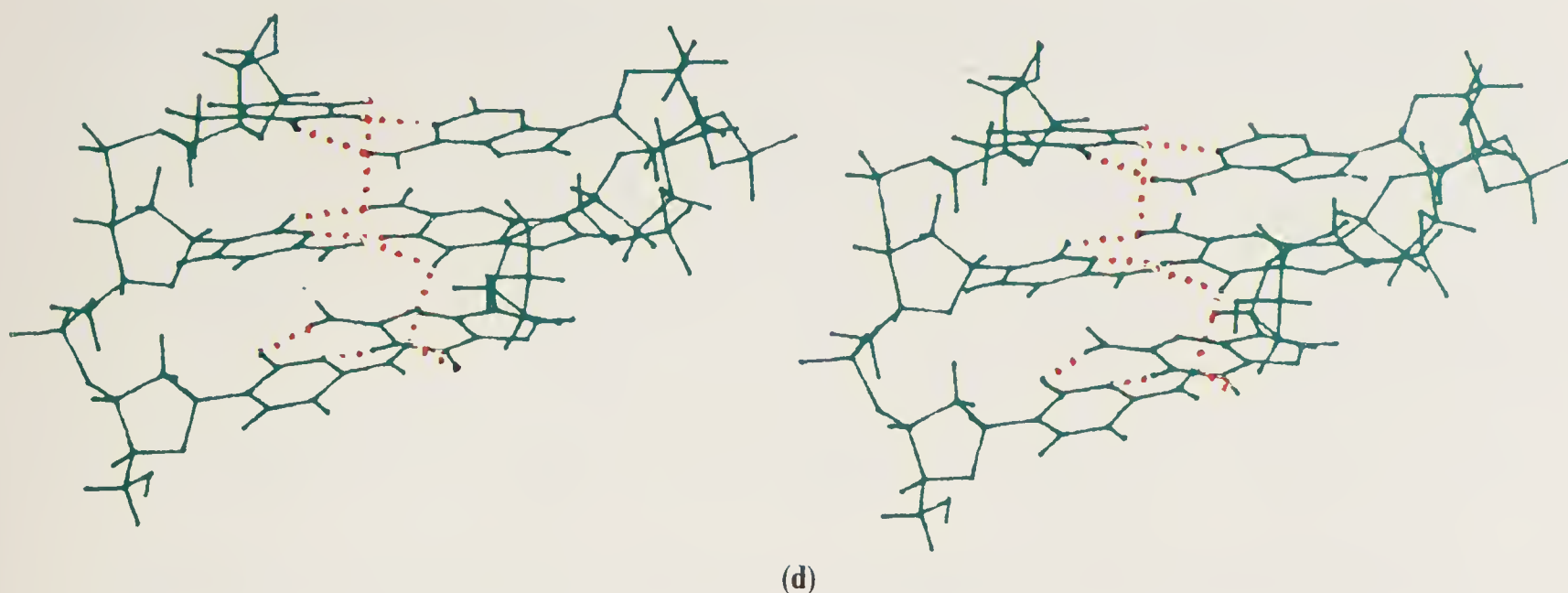


Figure 4. Stereo plots of a few representative energy minimized duplex trimers showing CSH at contiguous base steps (a) CSH at TG and GT steps in the sequence TGT/ACA. Note the central GC basepair taking a low propeller twist compared to flanking TA base pairs (b) CSH at GT and TG step in the sequence GTG/CAC. Note N1 (G) forms a bifurcated CSH with N1(A). Also note the difference in propeller twists of GC and TA basepairs at the TG step. (c) CSHs in the sequence CCA/TGG. The moderate propeller twists in TA and CG base pairs can be observed. (d) CSHs in the sequence CCT/AGG. The central G base participates in both the minor and major groove CSHs.

be inherent features of DNA helices especially of the preponderant B-type. While this study was being completed, crystal structures (13, 14) of a few oligodeoxynucleotides reported the occurrence of all the major groove cross strand hydrogen bonds confirming the predictions from model studies (4, 5). The initial observation (2) of a CSH between A and T has now been found as a recurring feature in several structures comprising AT sequences (15–17). While a G...T minor groove hydrogen bond has been found in a structure (18) possessing a mismatch, other suggested possibilities of minor groove CSHs are yet to be observed experimentally. It is hoped that these and other CSHs suggested from the present studies will find experimental evidence as more crystal structures are determined and analysed. A G...A CSH is the only one observed in a A-type helix (19). A bifurcated hydrogen bond involving N1(G) and O2(C) within a G–C base pair has been noted recently (20).

Sequence Dependent Propeller Twists in B-DNA Helices

An indirect outcome of the above calculations on dimer and trimer duplexes is the revelation that the magnitude of propeller twists are not uniform but assume very different values in the range -20.5° to -3.4° (table 4) depending on the neighbouring base pair sequences. An attempt is made here to obtain a possible relationship between the magnitude and the sequence. The dependence of base pair propeller twists on its neighbours is quite evident, for example from the large range (-20.5° to -3.4°) of propeller twists assumed by the GC base pair (table 4). High values ($\simeq -20^\circ$) are found in the homopolymeric CC/GG sequence and very low values in the sequences TG/CA (-11°) and GT/AC (-1°). It is intriguing to examine the causes for the low value of propeller twist found for the G–C pair. It is observed from consideration of model studies that the preference for the low value is to avoid unfavorable steric

Table 4. Propeller twists and CSH geometry in energy minimized duplex trimers. Arrows have same meaning as in table 1.



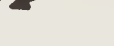


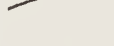


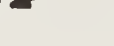





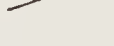





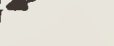


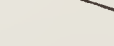





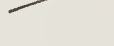






Trimer Sequences	Propeller Twist (°)	Bifurcated CSH-Bond Geometry		
		O ... H (Å)	N ... N (Å)	Angle (°)
T  A	− 13.00	3.30	3.50	98.6
T  A	− 9.62	3.19	3.47	101.2
T  A	− 19.19			
C  G	− 7.95	2.60	3.19	118.4
C  G	− 11.81	2.25	2.97	127.4
C  G	− 19.66			
T  A	− 9.66	2.74	2.97	93.4
G  C	− 3.38	3.10	3.70	118.9
T  A	− 16.45			
G  C	− 12.61	3.06	3.70	122.6
T  A	− 7.52	3.07	3.49	102.0
T  A	− 19.54			
T  A	− 13.04	2.97	2.98	80.5
C  G	− 11.39	2.23	3.00	128.1
C  G	− 20.55			
T  A	− 15.35	2.88	3.20	102.1
T  A	− 11.11	3.03	3.02	79.8
C  G	− 18.85			
T  A	− 14.04	2.75	3.30	103.8
T  A	− 8.29	2.72	2.92	91.3
G  C	− 3.75			
G  C	− 14.26	2.94	3.58	123.3
T  A	− 7.69	3.02	3.05	82.5
C  G	− 19.00			
G  C	− 16.38	2.78	3.40	124.1
T  A	− 5.45	2.73	2.93	91.1
G  C	− 3.17			
A  T	− 14.56	3.14	3.70	118.9
C  G	− 8.24	2.57	3.20	125.3
C  G	− 19.12			
T  A	− 12.74	3.07	3.09	81.5
C  G	− 6.19	2.65	2.89	93.6
A  T	− 9.56			
C  G	− 8.84	2.60	3.10	118.7
C  G	− 6.82	2.63	2.87	93.5
A  T	− 8.93			

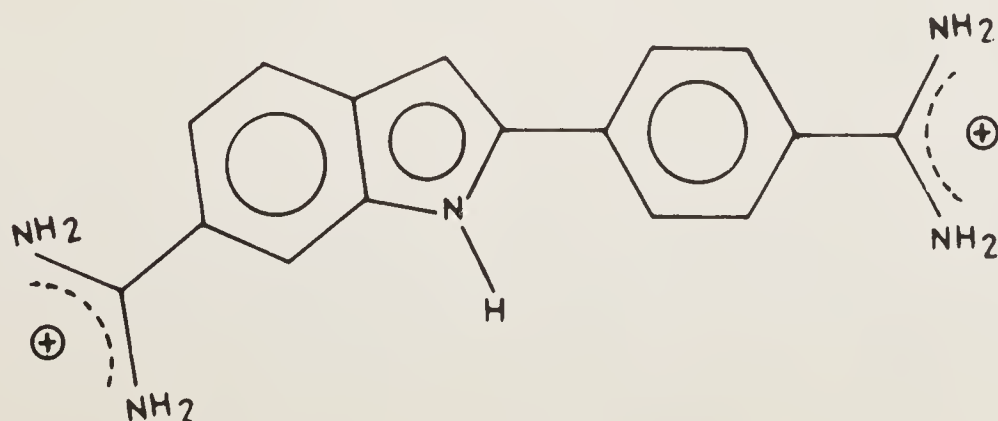
Table 5(a). Comparison of propeller twists (in degrees) of TA and GC base pairs at TG steps with X-ray data. Note that a pyrimidine-purine base pair follows them.

Base step	Reference	This Work		(14)		(2)		(10)		(12)		(2)	
TG	Top base pair	T	A	C	G	C	G	T	A	C	G	C	G
GC		− 3.38		− 5.2		− 7.7		− 6.0		− 11.0		− 11.0	
TA		− 16.45		− 15.5		− 14.6		− 17.0		− 25.0		− 19.0	
5'													

Table 5(b). Comparison of propeller twists (in degrees) of TA and GC base pairs at TG steps with X-ray data. Note that a purine-pyrimidine base pair follows them.

Base step	Reference	This Work		(11)		(10)		(15)	
TG	Top base pair	G	C	G	C	G	C	G	C
GC		− 8.24		− 9.0		− 13.0		− 9.6	
TA		− 14.56		− 8.0		− 14.0		− 11.0	
5'									

interactions between the amino group of guanine and C2 of adenine in the minor groove in the sequence TG/CA and the thymine methyl group and imidazole of guanine in the sequence GT/AC. They can be overcome in the TG/CA step by assuming moderate values of propeller twists either in both the base pairs or one of them assuming a lower value. On the other hand, it is necessary for the G–C pair to assume lower values of propeller twist in the sequence GT/AC. This correlation found in dimer duplexes is retained for the GT steps even in the trimer sequences (table 4) TGT/ACA (-3.4°), GTG/CAC (-3.2°), GTT/AAC (-3.7°), ACT/AGT (-6.2°) and ACC/GGT (-6.8°). This trend of low values of propeller twist for the G–C pair at GT steps compares very well with the experimentally observed values of -6° (GC) and -14° (AT) found in the crystal structure of a dodecamer (13). G–C pair takes a lower value of propeller twist even at the TG steps in the sequences TGT/ACA (-3.4°) and TGG/CCA (-8.2°) (table 4). On the other hand, the value is higher in the sequences TTG/CAA (-12.6°), CTG/CAG (-14.3°) and GTG/CAC (-16.4°). This could be due to its occurrence at the terminus in these sequences. Thus, the propeller twist of GC at TG step seems to be strongly influenced by the

**Figure 5.** Structure of DAPI

occurrence of a base pair (GC or TA) at the 3' end. Examples of the G–C pair taking up lower values (table 5a) of propeller twist at TG steps are seen in the available X-ray structure data (2, 14–18) in excellent agreement with the suggested sequence dependent correlation of propeller twist. The propeller twist of GC at TG step when it is flanked by a G–C pair at 3' end is compared with a few experimental observations (table 5b). It is seen that both the base pairs take up moderate values of propeller twist of nearly equal magnitude in contrast with the values obtained from this calculation. This might be due to the influence of the base pair occurring at the 5' end of the TG step, as is the situation in crystal structures. Hence, it is necessary to analyse further, tetrameric sequences or even longer ones to delineate fully detailed information on the base sequence dependence of propeller twist of this sequence as well as other sequences in general. Nevertheless, from the present studies, the magnitude of the propeller twists is clearly sequence dependent. Other recurring variations are also seen in table 4, seemingly suggestive of other sequence dependent correlation of propeller twists. More detailed calculations on a few of these sequences are required to substantiate them. Understanding of these sequence dependent properties will greatly facilitate correlation of conformational heterogeneity with base sequences in DNA double helices.

DAPI–DNA Interactions

DAPI (4',6'-diamidino-2-phenyl indole) (figure 5) is a cationic dye found to interact with DNA predominantly in a non intercalative mode but specifically with AT, but not GC rich regions. Solution studies (21, 22) show that its binding properties mimic netropsin despite the lack of extended chain-like chemical structure. A stereochemical study of interaction of DAPI with DNA has been taken up to examine the different modes of binding, types of interactions responsible for the stability of DNA–DAPI complex, features responsible for specificity and possible drug-induced conformational variations in DNA. Interaction energies of DAPI with a septamer duplex in the B-DNA geometry have been calculated for different base sequences to mimic poly(dA).poly(dT) and poly(dG).poly(dC). While it is known that DAPI spans not more than 3 base pairs, a septamer has been chosen to mimic minor groove of DNA and also to examine the overall effect of binding of DAPI on DNA helix. DNA–DAPI complex is energy minimized taking into consideration variations in the conformations of DNA as well as DAPI.

Since DAPI is asymmetric with respect to the central imino proton, it was felt important to consider the binding of DAPI in both the orientations related by 180° to determine the energetically preferred mode of binding. When the indole ring of DAPI is towards the 5' end of the pyrimidine strand it is referred to as mode A binding and when it is towards the 3' end it is referred to as mode B. These may correspond to two distinct binding sites seen in experimental studies (22). An earlier study (23) not only considered a fixed DNA geometry but also a single orientation of DAPI. Different starting conformations corresponding to different positions of DAPI in the minor groove of B-DNA helix are considered to arrive at the energetically preferred mode of binding. Interestingly, it is found that DAPI in all cases favors an orientation facilitating hydrogen bonding of the imino hydrogen with O2 of thymine. Stereo views of the energetically favored binding of DAPI in both A and B modes are shown in figures 6 and 7 respectively. In mode A binding, indole guanidinium is seen

to form strong hydrogen bonds with O2 of successive thymines while the phenyl guanidinium is involved in a hydrogen bond with phosphate of adenine strand. The central imino hydrogen forms a bifurcated hydrogen bond with both O2 and N3 involving bases across. On the other hand, in mode B binding (figure 7), only guanidinium of indole forms a hydrogen bond with O2. The central indole hydrogen is engaged in a strong hydrogen bond with O2 and a weak one with N3 across the strand and one base pair step lower than that of the mode A binding. However, in both modes of binding, DAPI is found to occur nearly along the centre of groove sandwiched between the sugar–phosphate backbone chains. The phenyl ring occurs a bit deep inside the groove in mode B compared to mode A. Because of this mode B is found to be energetically more stable by about 10 kcal/mole arising mainly through electrostatic contributions. From the point of view of possible specificity arising through hydrogen bonds, mode A interaction is preferred. In fact this mode of binding is seen in X-ray crystal structure (24) of a dodecamer complexed with DAPI.

In order to account for the factors that are responsible for the specificity of DAPI to AT-rich regions, interaction energies of DAPI with (dG)₇·(dC)₇ have also been

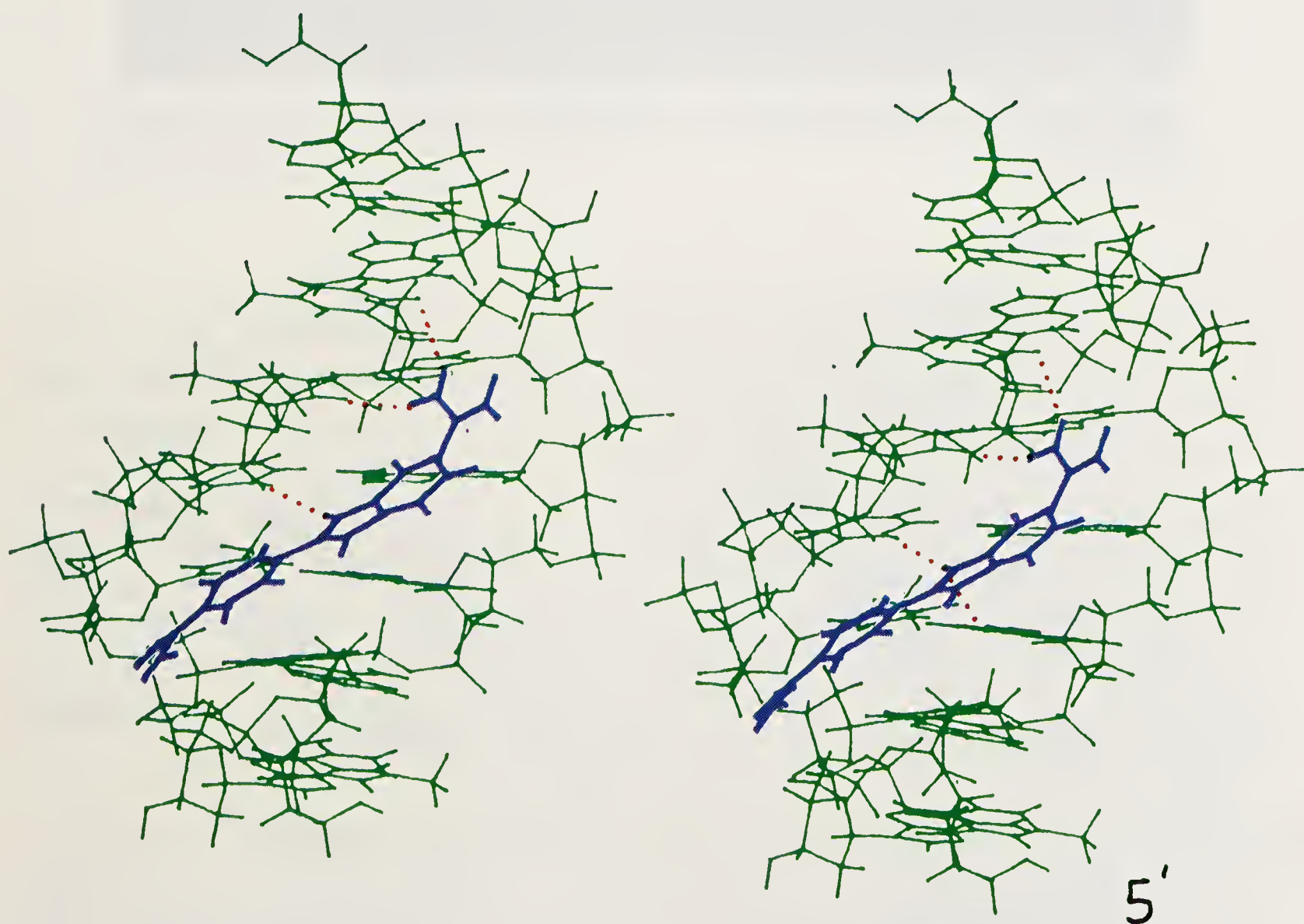


Figure 6(a). Stereo view of mode A binding of DAPI to (dA)₇·(dT)₇. Hydrogen bonds between DAPI and DNA are shown by dotted lines. The central imino nitrogen forms a bifurcated hydrogen bond with O2 of thymine and N3 of adenine across. Phenyl guanidinium hydrogen bonds with nearby phosphate while rotated indole guanidinium forms two hydrogen bonds with O2 of adjacent thymines. The indole ring is stacked between sugar rings. Note the curving of DAPI along the groove.



Figure 6(b). Space filling representation of figure 6(a) obtained using the inhouse program (28).

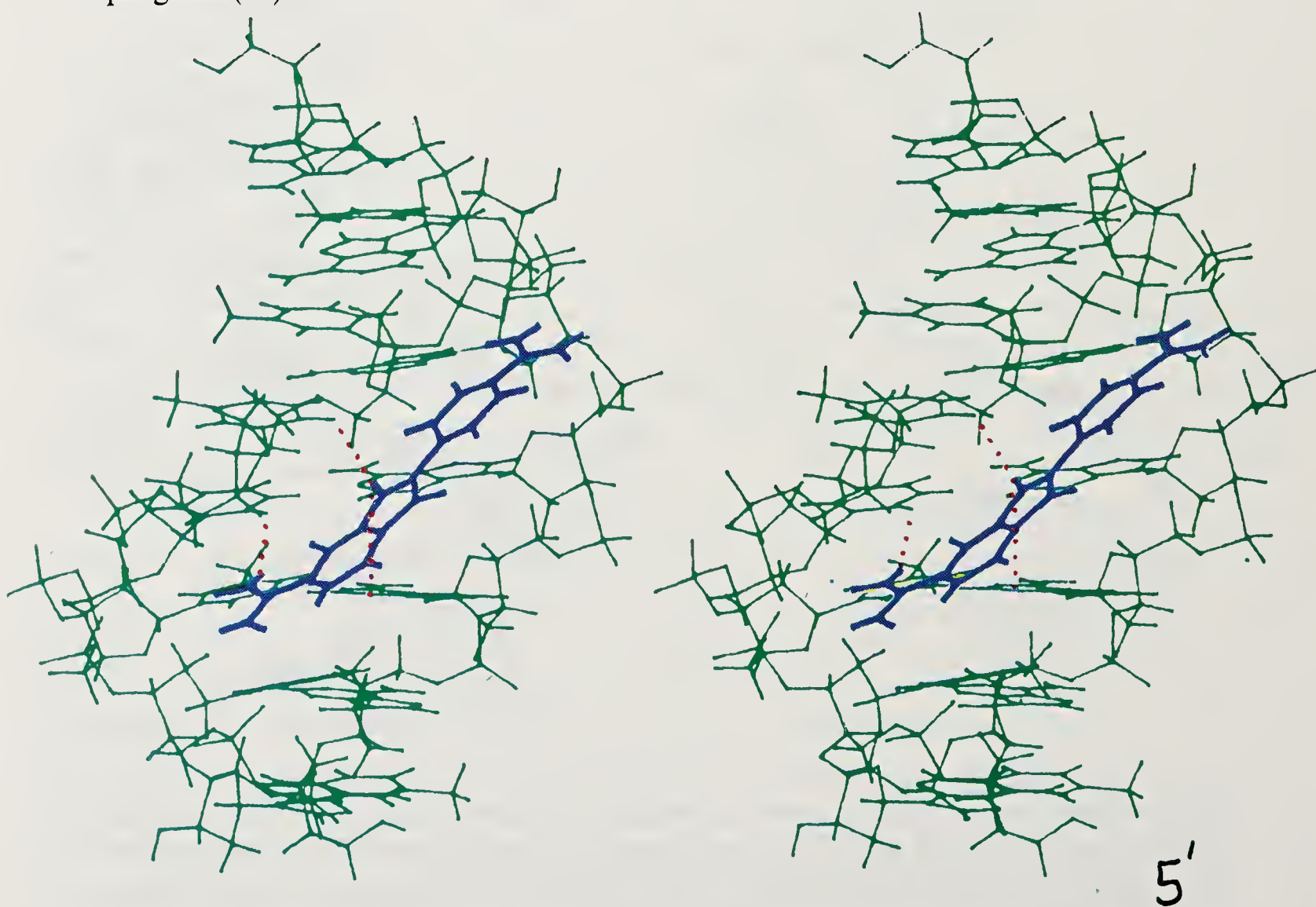


Figure 7. Stereo view of mode B binding of DAPI to (dA)₇.(dT)₇. The phenyl ring is located deep in the minor groove. Both the ring systems stack well between the sugar-phosphate backbone in the minor groove. Strong hydrogen bonds between indole nitrogen and O2(T) and a weak one with N3(A) are seen. Note the curving of DAPI along the groove.

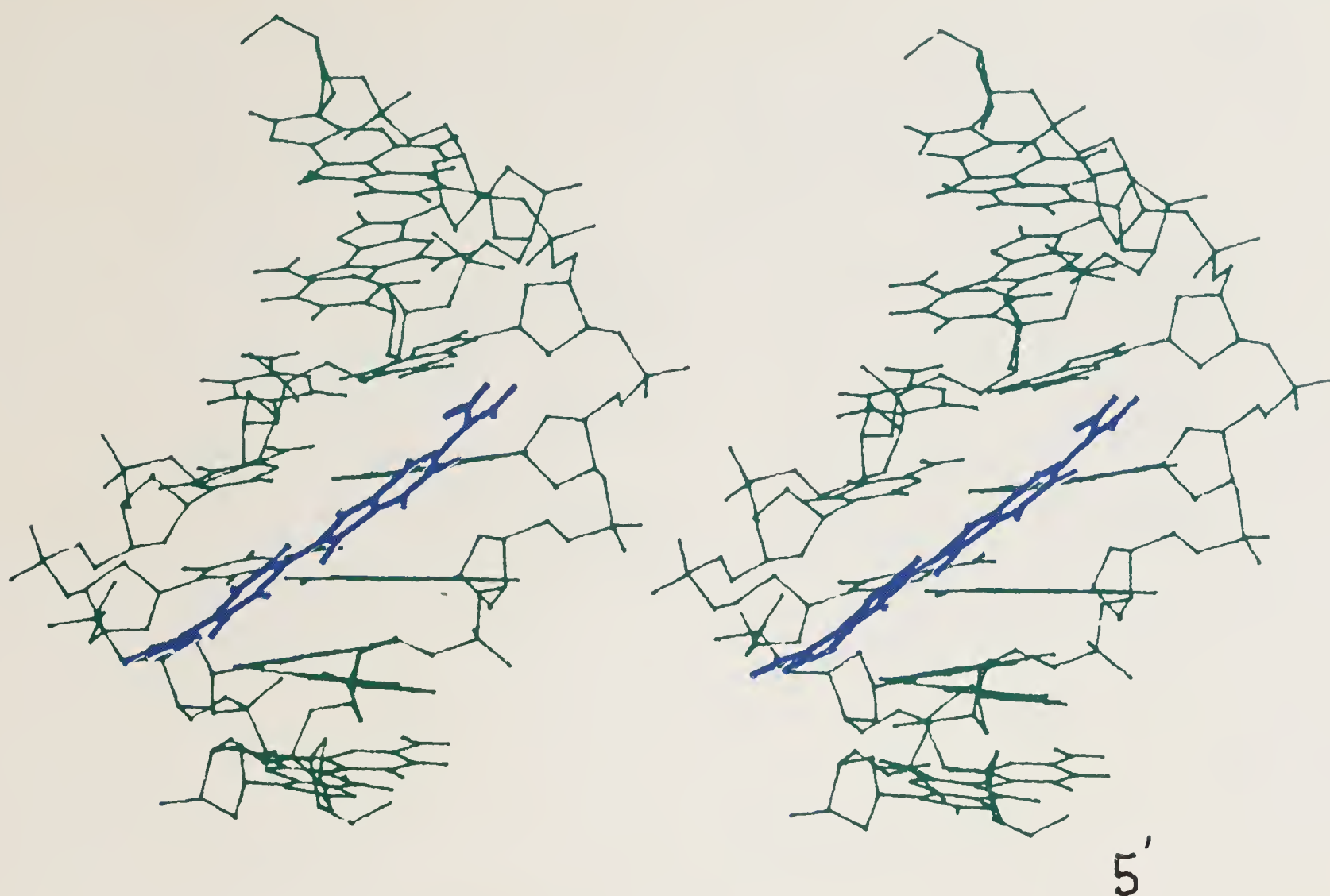


Figure 8. Mode B binding of DAPI to the sequence $(dG)_7.(dC)_7$. DAPI remains more or less planar and drifts outwards from the minor groove with no hydrogen bond being formed.

calculated. It is seen that DAPI is located (figure 8) in the minor groove of $(dG)_7.(dC)_7$ nearly in the same region as found for $\text{poly}(dA)\cdot\text{poly}(dT)$ with a slight outward displacement of especially the phenyl ring. Because of this no hydrogen bonds are found feasible. However, the six-membered ring of indole is flanked on either side by the sugar ring oxygens enhancing stacking interactions. Consequently van der Waals stabilisation energy here is even better than in $(dA)_7.(dT)_7$. However, the total energy is very high for both A and B modes due to the unfavorable electrostatic contributions which render the complex to be highly unstable compared to that with $(dA)_7.(dT)_7$. It should be pointed out that high energy is an intrinsic feature of the minor groove of the B-type structure of $(dG)_7.(dC)_7$ and not due to binding. This is consistent with the observation (25) that $\text{poly}(dA)\cdot\text{poly}(dT)$ possesses a large attractive molecular electrostatic potential in the minor groove. Nearly similar location of DAPI in the minor groove of $(dG)_7.(dC)_7$ and $(dA)_7.(dT)_7$ can be better seen in figure 9.

Possible binding of two molecules of DAPI simultaneously in the minor groove of $(dA)_7.(dT)_7$ is shown in figure 10. This is relevant since CD studies (26) suggest possible interactions between two close lying DAPI molecules in DAPI–DNA complex. One of the molecules is in the A mode and the other in B mode. It is found that the phenyl rings are located deep inside the minor groove while the indole rings are towards the periphery. In addition to stacking between two DAPI molecules, indole rings of DAPI are well stacked with sugar rings. The hydrogen bonding pattern here is different from those seen earlier with binding of a single molecule of DAPI. The imino nitrogens of two DAPI are engaged in single strong hydrogen bond with O2 and N3 of the same base pair. The minor groove widens to nearly 14 Å at the centre

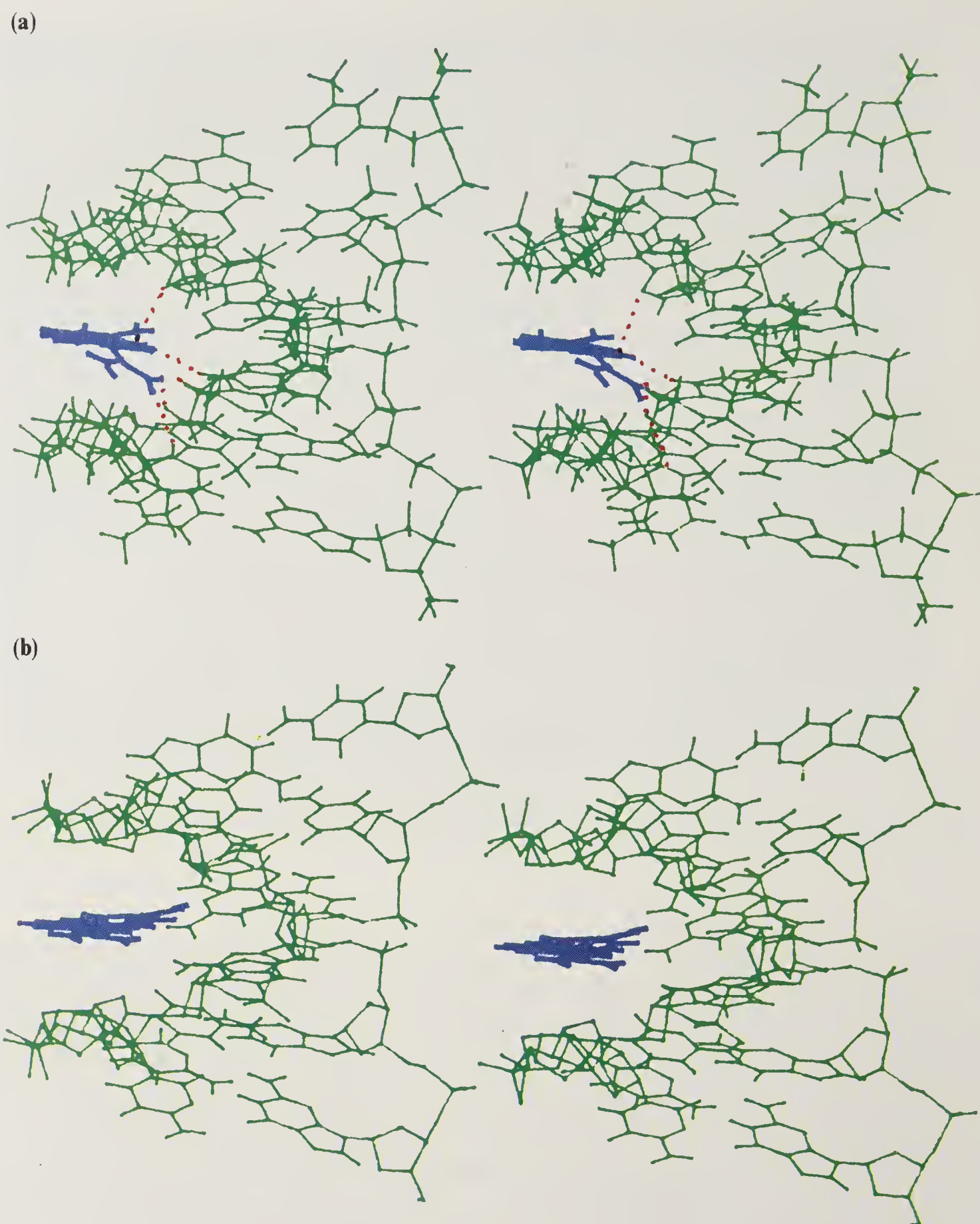


Figure 9. Location of DAPI in the minor groove viewed end on (a). $(dA)_7 \cdot (dT)_7$ (b) $(dG)_7 \cdot (dC)_7$.

to accommodate two DAPI molecules bringing out another facet of flexible nature of grooves of B-DNA helices.

Conclusion

It is clear from above that cross strand hydrogen bonds in DNA helices need not be restricted to homopolymeric sequences but they can be realised also in a few other limited sequence combinations on both major groove and minor groove sides. While

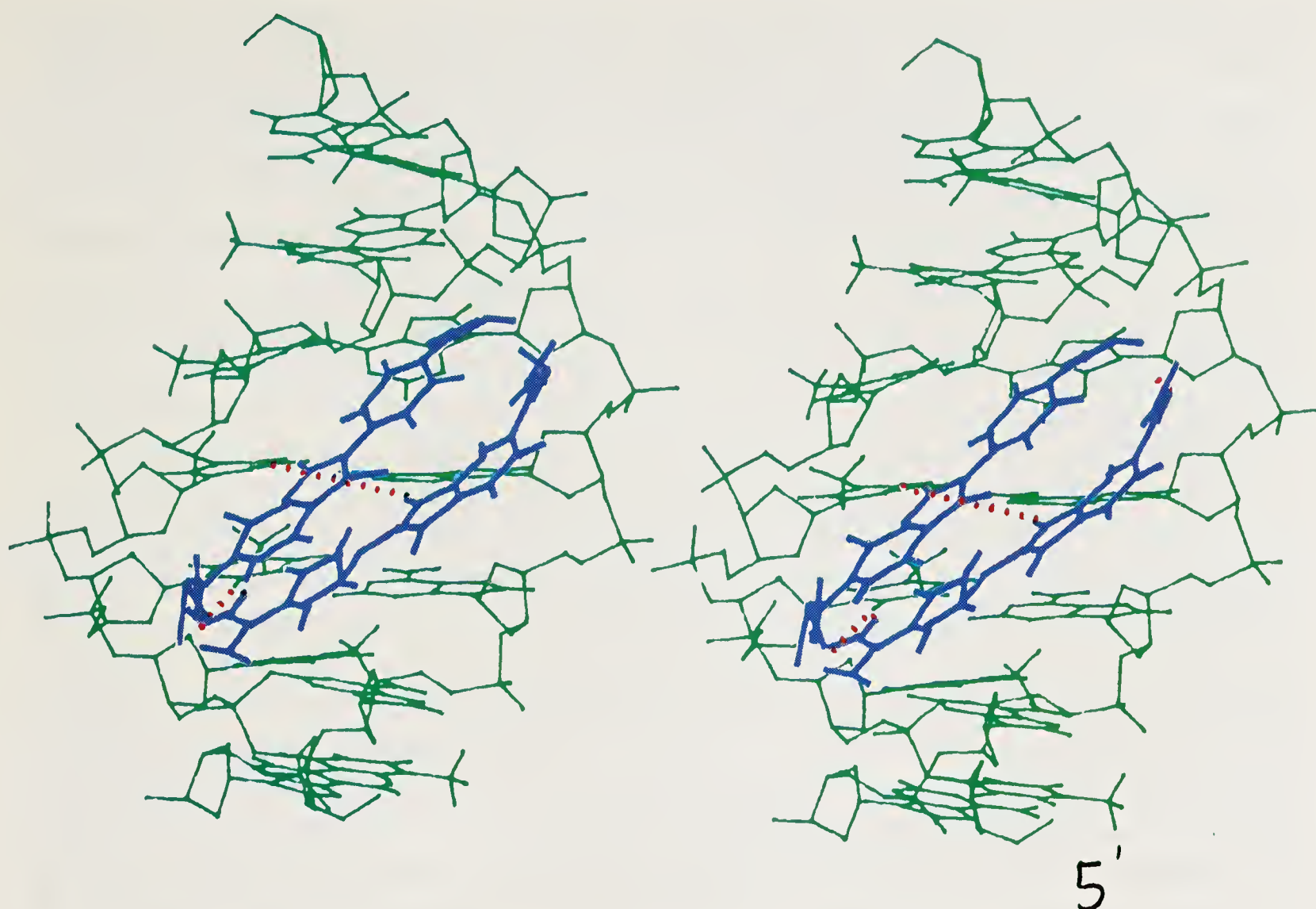


Figure 10(a). Two DAPI molecules one in mode A and the other in mode B bound in the minor groove of (dA)₇·(dT)₇. Mode A molecule forms two strong hydrogen bonds involving its indole nitrogen and phenyl guanidinium with adenine N3 and thymine O2 respectively. Mode B molecule forms only one hydrogen bond between its indole nitrogen and thymine O2. Widening (≈ 14 Å) of the minor groove at the center is seen.

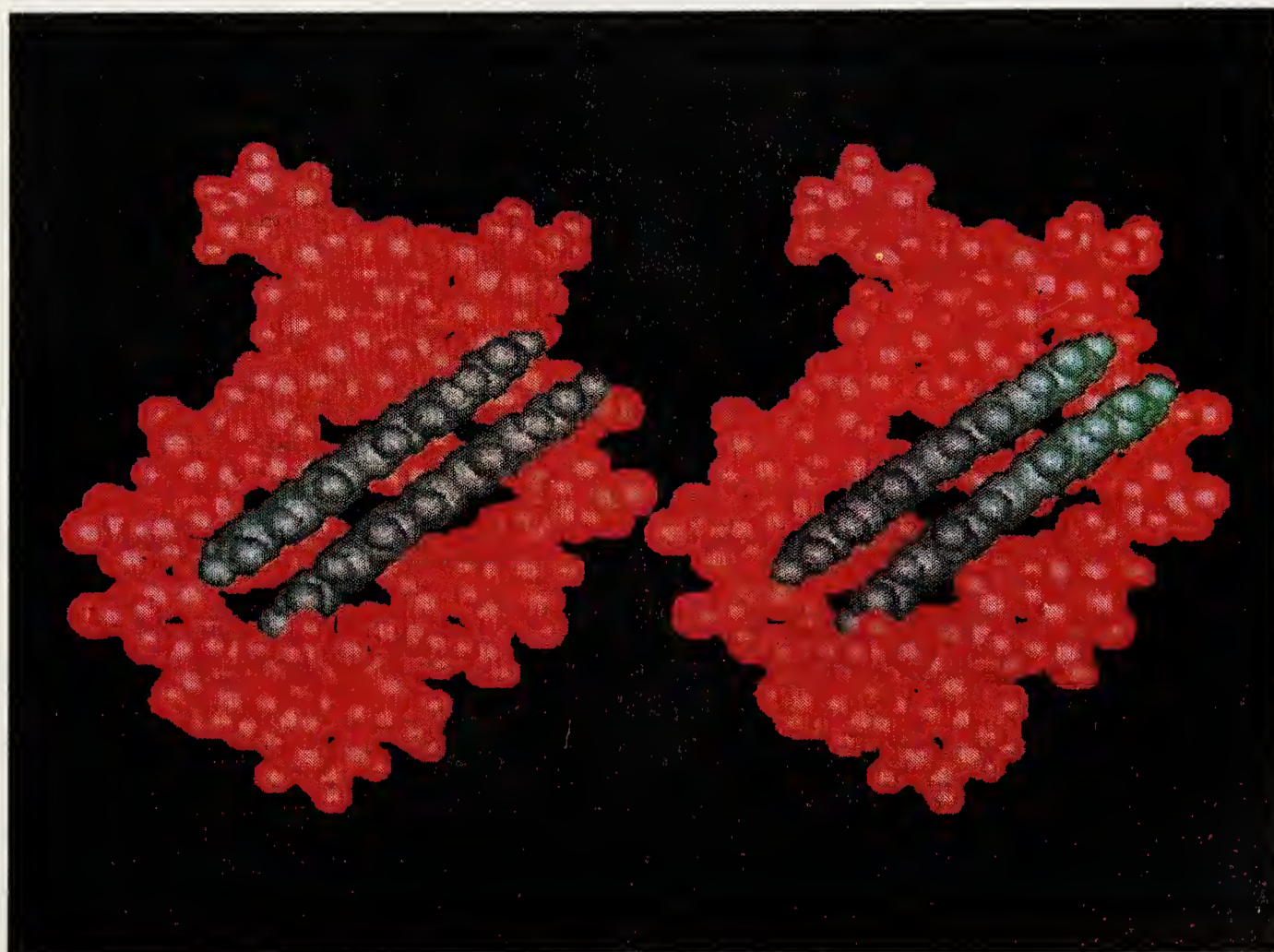


Figure 10(b). Space filling representation of figure 10a.

their occurrence depends to some extent on the helix type, the propeller twist driven CSHs are to be regarded as yet another general feature intrinsic to DNA double helices especially in the biologically significant B-type. Since our predictions (4–6), nearly all of them have been observed in X-ray structures containing appropriate sequences. It is found that cross strand hydrogen bond involving purine...pyrimidine in homopolymeric sequences seem quite natural and strong while those between purine...purine, pyrimidine...pyrimidine require some modification in base pair parameters compared to canonical B-DNA. Thus, occurrence of CSH in A-tracts and G-tracts may endow them with distinct features different from bulk DNA as has been seen already in certain A-tract containing structures (2). It is therefore reasonable to suggest that G-tracts may also exhibit properties similar to A-tracts especially in relation to DNA bending. Further studies especially crystal structure investigations on designed oligomers promoting CSH (other than homopolymeric) should elucidate if these contribute to the characteristic properties exhibited by homo- and alternating polynucleotides. Similarly, the role of other bifurcated hydrogen bonds (cross-strand or otherwise) in relation to non-uniform DNA conformation needs to be evaluated. One thing that is certain is that the proton deficient nucleic acid bases tend to optimise their associations horizontally and vertically through maximal hydrogen bonding interactions. An off shoot of this analysis has been a neat correlation observed for the first time between propeller twist and base sequence in B-DNA helices. Results from interaction study of DAPI with DNA show that the presence of 2-amino group does not *a priori* prevent the binding of DAPI to poly(dG).poly(dC), but it is the charge effects revealed as unfavorable electrostatic interactions that are responsible for the preference of DAPI to poly(dA).poly(dT). It is shown that two stacked DAPI molecules can simultaneously bind to the minor groove of B-DNA in an energetically favoured mode concomitant with widening of minor groove by about 3 Å.

Acknowledgements

Part of this work was supported by a grant from Department of Science and Technology, Government of India. S. M. thanks U.G.C. for a fellowship. The authors thank Prof. P. Kollman for the AMBER program.

References

1. Dickerson, R. E. (1983) *J. Mol. Biol.* **166**, 419–441.
2. Nelson, H. C. M., Finch, J. T., Bonventura, F. L. & A. Klug (1987). *Nature (London)* **330**, 221–226.
3. Koo, H. S., Wu, H. M. & Crothers, D. M. (1986) *Nature (London)* **320**, 501–506.
4. Mohan, S. & Yathindra, N. (1988) *Book of Abstracts*, National Seminar on Crystallography, (Varanasi, India).
5. Mohan, S. & Yathindra, N. (1989) *Book of Abstracts*, Sixth Conversation in Biomolecular Structure and Dynamics, (Albany, N.Y., U.S.A.)
6. Mohan S. & Yathindra, N. (1990) *Book of Abstracts*, Indo-Soviet Workshop on Protein–DNA interactions and crosslinking (Bangalore, India).
7. Arnott, S. & Hukins, D. W. J. (1972) *Biophys. Biochem. Res. Commun.* **47**, 1504–1509.
8. Dickerson, R. E. (1989) *J. Biomol. Str. Dyn.* **6**, 627–634.
9. Parthasarathy, R. (1969) *Acta. Crystallogr.* **B25**, 509–518.

10. Jeffrey, G. A. (1989) in *Landolt-Börnstein*, New Series, Group VII, ed. Saenger, W. (Springer, Berlin, Heidelberg and New York), pp. 277–348.
11. Shakked, Z. & Rabinovich, D. (1986) *Prog. Biophys. Mol. Biol.* **47**, 159–195.
12. McCall, M., Brown, T. & Kennard, O. (1985) *J. Mol. Biol.* **183**, 385–396.
13. Timsit, Y., Westhof, E., Robert, P. P. F. & Moras, D. (1989) *Nature (London)* **341**, 459–462.
14. Heinemann, U. & Alings, C. (1989) *J. Mol. Biol.* **210**, 369–381.
15. M. Coll, Christin, A. F. Wang, A. H.-J. & Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8385–8389.
16. Yoon, C., Prive, G. G., Goodsell, D. S., & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6332–6336.
17. Gabriele, A. D., Sanders, M. R. & Steitz, T. A. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 1816–1820.
18. Prive, G. G., Kan, L. S., Kopka, M., Heinemann, U., Chandrasegaran, S. & Dickerson, R. E. (1987) *Science* **228**, 498–504.
19. Jain, S., Zon, G. & Sundaralingam, M. (1989) *Biochemistry* **28**, 2360–2364.
20. Takusagawa, F. (1990) *J. Biomol. Str. Dyn.* **7**, 795–808.
21. Kapuscinski, J. & Szer, W. (1979) *Nucleic Acids Res.* **6**, 3519–3534.
22. Manzini, G., Barcellona, M. L., Avitabile, M. & Quadrifoglio, F. (1983) *Nucleic Acids Res.* **11**, 8861–8875.
23. Gresh, N. (1985) *Int. J. Biol. Macromol.* **7**, 199–202.
24. Larsen, T. A., Goodsell, D. S., Cascio, D., Grzeskowiak, K. & Dickerson, R. E. (1989) *J. Biomol. Str. Dyn.* **7**, 477–491.
25. Pullman, A. & Pullman, B. (1981) *Stud. Biophys.* **86**, 95.
26. Kubista, M., Akerman, B. & Norden, B. (1987) *Biochemistry* **26**, 4545–4553.
27. Srinivasan, A. R. & Olson, W. K. (1987) *J. Biomol. Str. Dyn.* **4**, 895–938.
28. Balasubramanian, R. & Rafi, Z. A. (1989) *Biosystems* **23**, 15–19.

Solvent Accessibility of Nucleic Acids

P. K. Ponnuswamy

Department of Physics, Bharathidasan University, Tiruchirapalli 620 024,
Tamil Nadu, India

Several attempts have been made in the past to gain insight into the solvent structure around the biological molecules, and also on the forces operating in the solvent–solute interfaces. Such information will be of immense help in understanding the functional roles of macromolecules *in vivo*. One such notable study is that by the Richards group (1,2) which yields information on the solvent-accessible parts in macromolecular systems. Chothia (3,4) applied this technique to a number of protein molecules in crystalline state and interpreted the results to emphasize the role of hydrophobic forces in the stability of protein structures. An important outcome of this study was the correlation between the accessible surface area and the hydrophobicity in protein molecules. We were the first to apply this technique to nucleic acid systems (5–7). Starting from representative subunits, we studied the solvent accessible character of tRNA molecules in crystalline forms, and DNA segments in specified duplex states. This article briefly summarizes the salient results of this extended study made during the past several years.

The Method of Study

In the method followed, the solute molecule was represented by a set of interlocking spheres of appropriate van der Waals radii assigned to each atom, and the solvent molecule was rolled along the envelope of the van der Waals surface at planes conveniently sectioned, and the total surface accessible to the solvent molecule – and hence the solvent accessibility of various atoms of the solute molecule for different conformations – were computed. The surface area generated by the set of interlocking spheres centered at the corresponding atomic positions of the solute molecule in a particular conformation is known as the ‘van der Waals area’; when the solvent molecule is rolled over the van der Waals surface area of the solute molecule, the surface generated by all the loci of the center of the solvent molecule is known as the ‘accessible surface area’; the van der Waals surface area of the solute atom with which the solvent can have contact when placed over it without penetrating any other atom is called ‘the contact area’. Figure 1 shows a pictorial representation of these areas for a hypothetical solute system, water being the solvent molecule. The numerical values of these three parameters and an associated parameter, ‘the solvent accessibility’, S (the extent of free surface available for interaction with a solvent molecule) were computed as per the formulation of Richards (1,2).

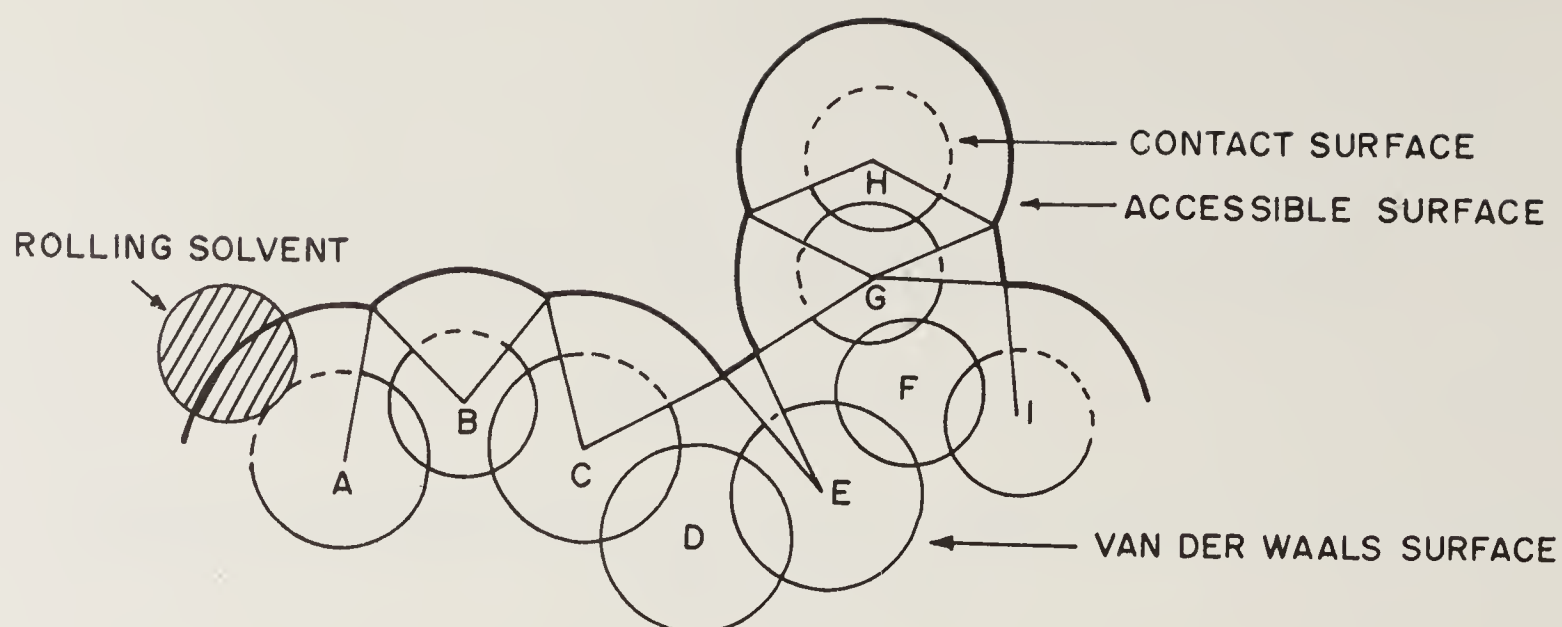


Figure 1. A cross-section of a part of a macromolecule in space. A, B, ...I are atoms. van der Waals, contact, and accessible surface areas (see text for definition) are indicated. Atoms D and F are inaccessible to the solvent molecule.

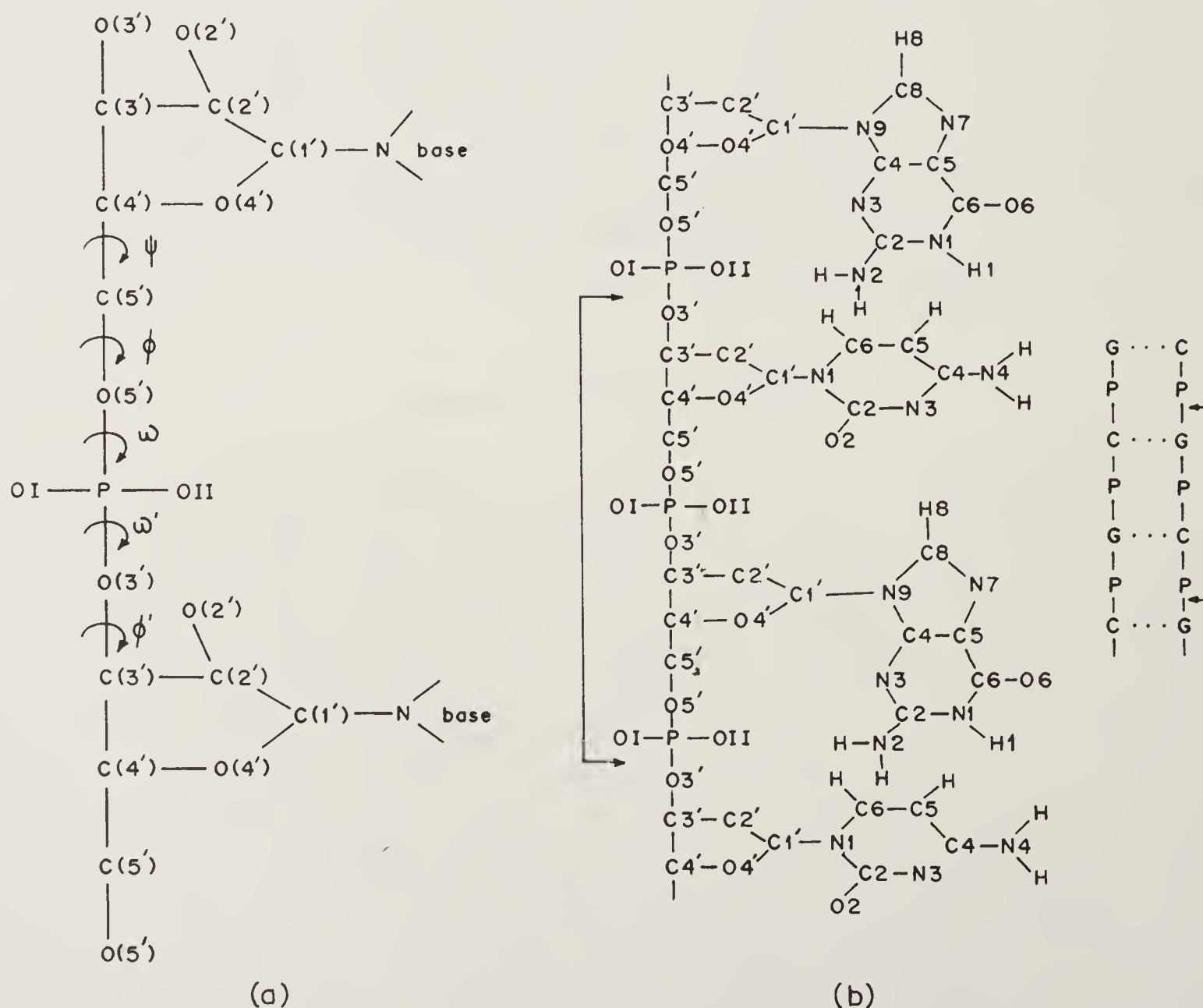


Figure 2. The skeleton, numbering convention and the dihedral angles of (a) an RNA dimeric subunit, and (b) a DNA tetranucleoside triphosphate subunit; in b, atoms included in the calculation are indicated within arrows; the double helical part considered is also depicted.

Results

Dimeric Subunits

To start with we considered a few representative dinucleosides belonging to RNA and DNA systems. The skeletal form of a dimeric unit of RNA is shown in figure 2a. Atomic solvent accessibilities were calculated for such molecular units of RNA and DNA as a function of the three dihedral angles, ω' , ω , and ψ , keeping ψ at positions 60° , 180° and -60° , and varying ω' , ω at the intervals of 20° each. The ϕ and ϕ' angles were kept at 180° and -135° respectively. Using the computed values we constructed solvation maps in the (ω', ω) space for each unit. Figure 3 is one such map corresponding to $\psi = 60^\circ$ for the O(2') atom of the ribofuranose at the 3' side

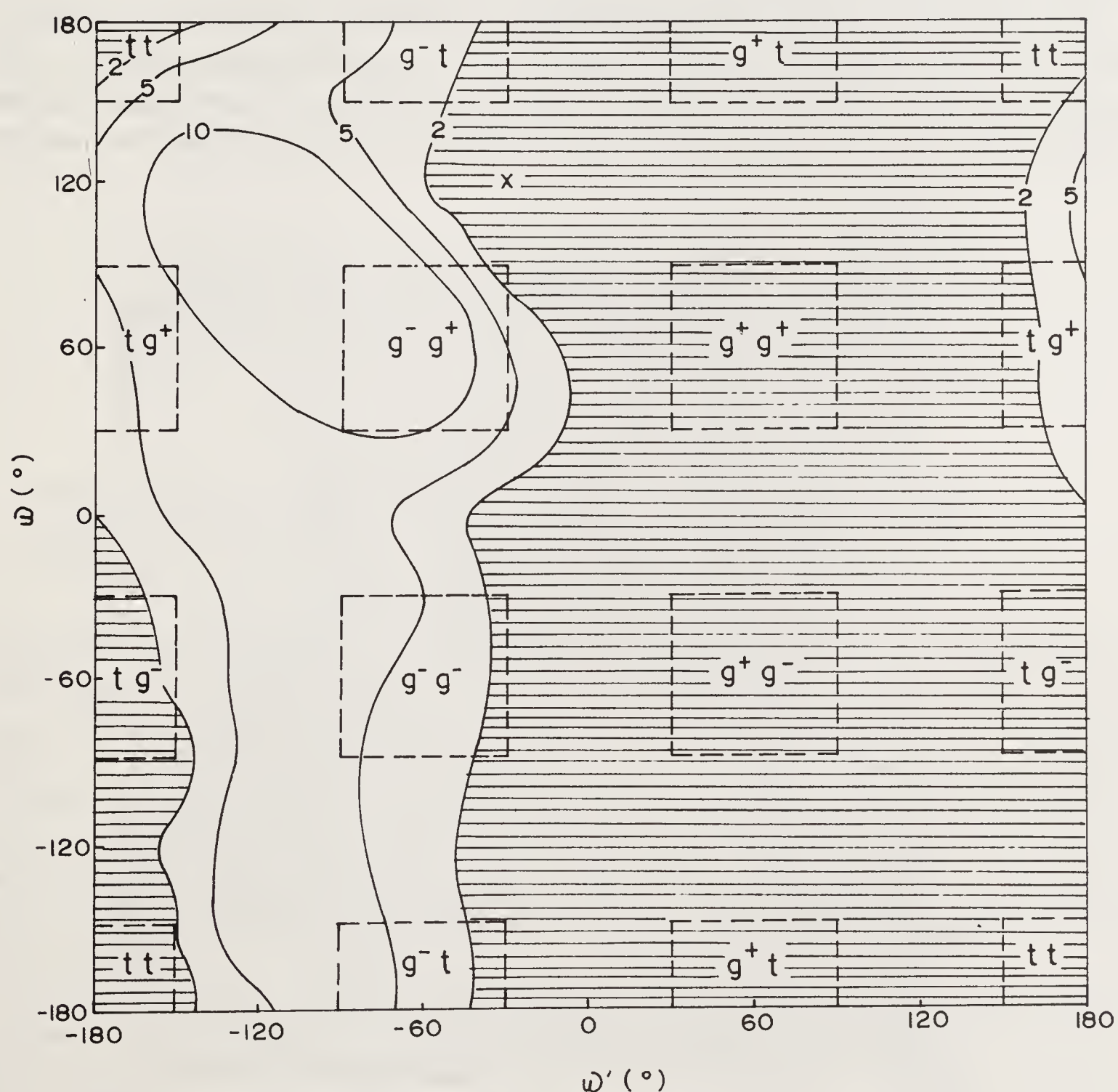


Figure 3. Solvent accessibility map for the hydroxyl O(2') atom of the furanose ring of the RNA subunit shown in figure 2a. The location of maximum accessibility is marked by X. The area enclosed by the values X and X - 2 is shaded and designated by the contour number 2, and numbers 5 and 10 have similar meanings. The boundaries of different conformational domains in the (ω', ω) space are marked by the symbols, g^+ , t, and g^- .

of the dinucleoside monophosphate (figure 2a). In this map, X denotes the location of maximum accessibility, and accessibility contours at intervals of -2 , -5 , and -10 with reference to X have been marked. We note from the map that the atom $O(2')$ could most easily associate with a water molecule at $(\omega', \omega) = (-30^\circ, 120^\circ)$, and the best probable region for water interaction could be taken as the area covered by the contour $X - 2$. However, for further discussion, the probable region for solvent interaction will be taken as the area covered by the $X - 5$ contour.

Accessibility maps similar to figure 3 were computed for the atoms $O(3')$, $O(4')$, $O(5')$, OI , and OII , and they were used to construct what is called 'the superposition map' which provides the combined information of the solvent accessibility character of a dimeric unit. The superposition map for the RNA subunit with $C(3')$ -endo sugar and $\psi = 60^\circ$ is shown in figure 4. This map identifies the regions having higher solvent accessibilities for the different atoms with reference to the nine local conformational domains in the (ω', ω) space.

The superposition map for the DNA dimeric subunit with $C(2')$ -endo sugar and $\psi = 60^\circ$ is shown in figure 5. In this map, we note that an appreciable area with $\omega' \simeq -60^\circ$ to 40° and $\omega \simeq -60^\circ$ to 60° permits all four oxygen atoms to have higher solvent accessibilities; this four-atom region is, in fact, comparatively larger than that predicted

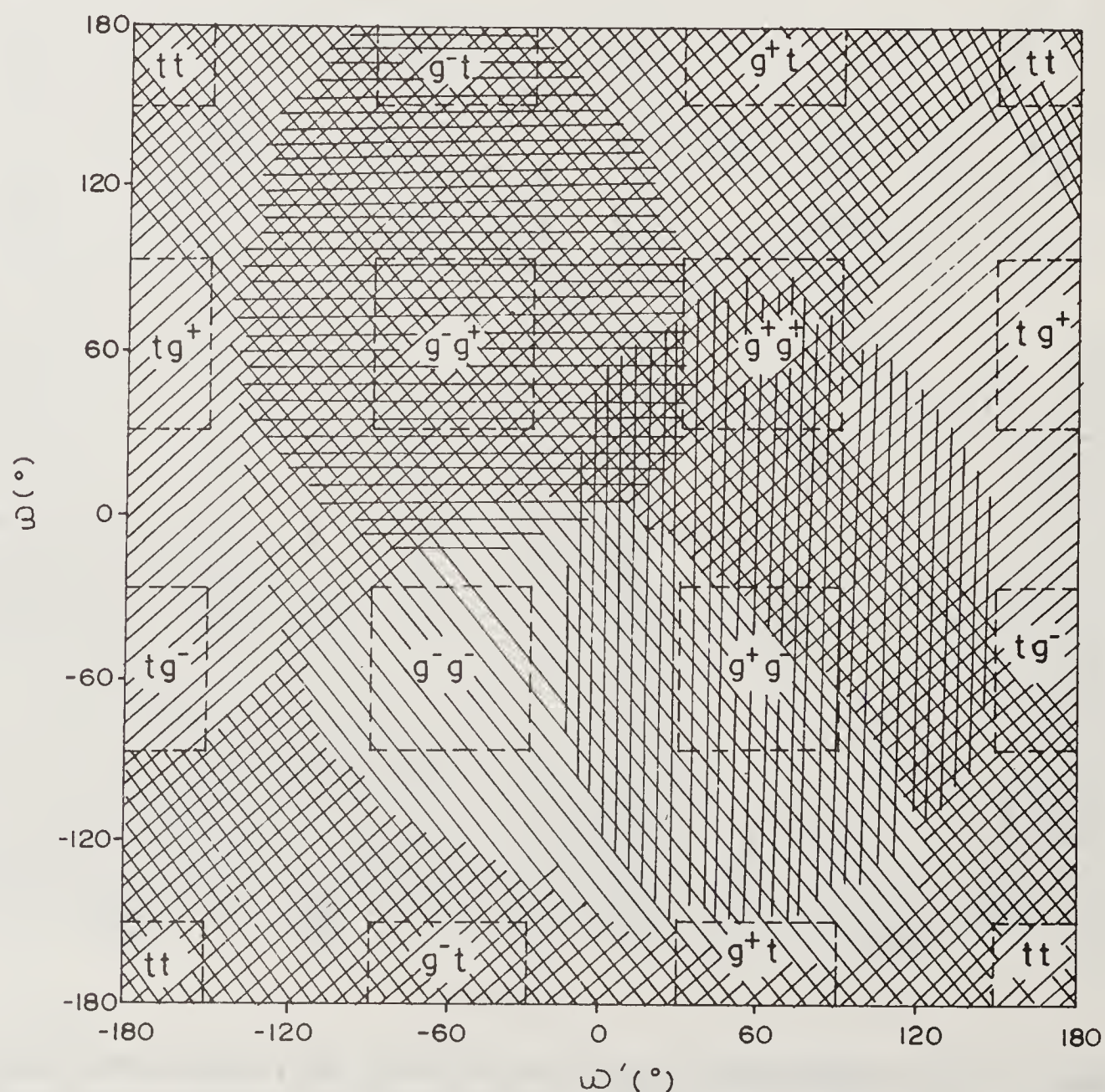


Figure 4. Superposition solvation map for the RNA dimeric subunit with $C(3')$ -endo sugar for $\psi = 60^\circ$. The shadings represent the areas of higher accessibilities for the oxygen atoms OI (▨), OII (▤), $O(3')$ (▥), and $O(4')$ (▧).

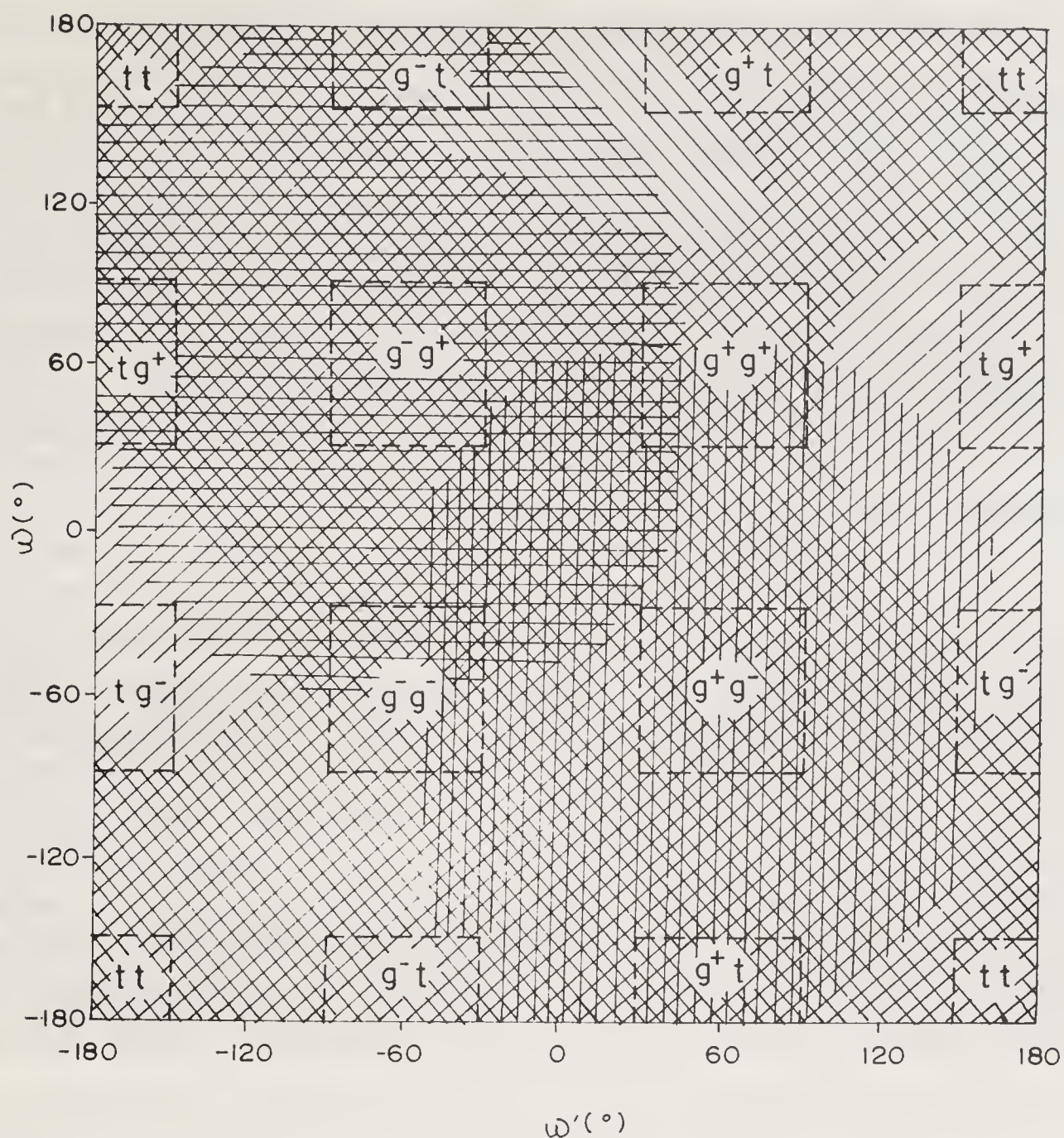


Figure 5. Superposition solvation map for the DNA dimeric subunit with C(2')-endo sugar for $\psi = 60^\circ$. Shadings indicate the same oxygen atoms as in figure 4.

for the corresponding case of $\psi = 60^\circ$ in the C(3')-endo subunit. An interesting aspect of this map is that the location where the B-DNA conformer ($\omega', \omega = -102^\circ, -41^\circ$) occurs permits higher accessibility to the three atoms OI, O(3'), and O(4'), while the location at which the C-DNA conformer ($\omega', \omega = -148^\circ, -45^\circ$) appears permits only the O(4') atom. Thus, in general, the conformation corresponding to the B-DNA structure exposes three backbone oxygen atoms to maximal solvent association, whereas those corresponding to the A- and C-forms expose only one atom, thus giving a clue to the more stable nature, and hence the dominant occurrence of the B-form of DNA.

Superposition maps (similar to figures 4 and 5) were constructed for the possible different kinds of dimeric subunits of DNA and RNA, and an analysis of them provided the following major results:

- 1) In general, the solvatable properties of the five backbone atoms OI, OII, O(3'), O(4'), and O(5') are more or less the same in the subunits of both RNA and DNA;
- 2) For both RNA and DNA subunits, the $\omega = trans$ orientation is more favorable to solvent association than is the $\omega' = trans$ orientation;
- 3) In case of the DNA subunits, the B-form conformer generally exposes more oxygen atoms to solvent than the A- and C-form conformers.

The $tRNA^{Phe}$ Molecule

As a next step, we made a study of the $tRNA^{Phe}$ molecule for which the atomic coordinates were then available from X-ray investigations (8,9). The computed accessibility values for the backbone polar atoms are plotted in figure 6. From this plot we note that the three atoms OI, OII, and O(2') placed throughout the chain have high solvent accessibility values – in the range of 20–45. This range is quite sufficient to envisage interactions in the forms of solute–solvent and solute–solvent–solute, as suggested by Bolton and Kearns from NMR studies (10). It is interesting to note from the variation of accessibility of the two phosphate oxygens along the entire stretch of the molecular chain that the OII atom has comparatively higher solvent exposure than OI at all the residues except at residues 7, 17, 22, 47, 49, and 59; hence, it is more exposed to solvent than OI. Only two residues, 50, and 60, orient themselves so as to restrict the phosphate oxygens from having contact with solvent. Similarly, only the residues 7–9, 21, 46, and 60 orient their sugar rings so that the O(2') atom is not in contact with solvent; these residues, most probably, have solute–solute hydrogen bonds involving the O(2') atom. The solvent accessibilities of the O(3') and O(4') of a number of residues lie between 5 and 10, indicating their restricted role in solvent–solute interactions. It is interesting to point out here that the O(5') atoms of most of the residues have very small solvent accessibilities (1.0 to 5.0) compared to the values of the corresponding O(3') atoms. Also, unlike those residues in the helical regions having uniform solvent accessibilities, the values of the residues in the various loop regions vary. This is due to varying nature of tertiary interactions present in the case of loop sections compared to those of the helical regions.

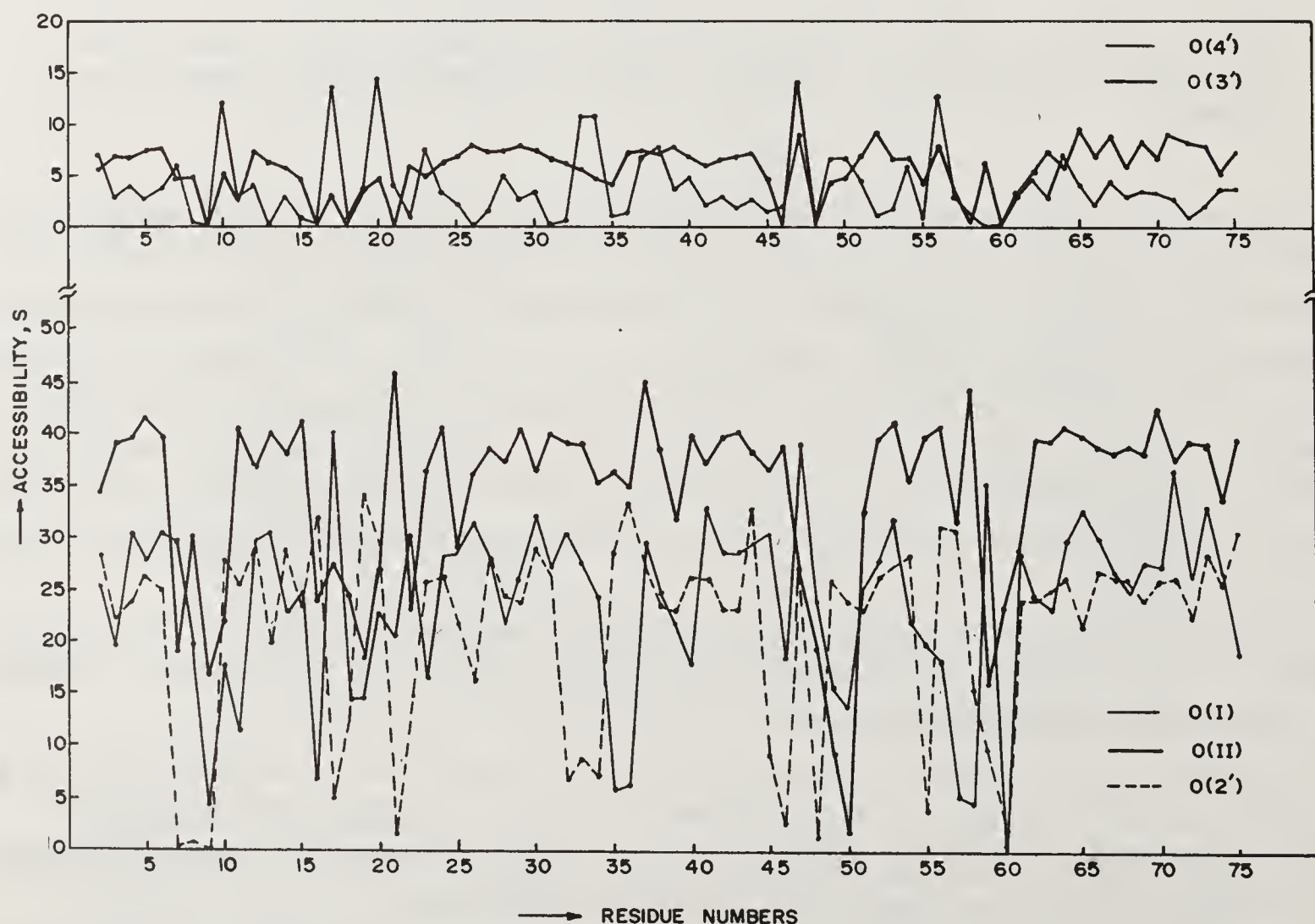


Figure 6. Plot of the accessibility values for the backbone atoms OI, OII, O(2'), O(3'), and O(4') of the residues of $tRNA^{Phe}$ in the crystalline form.

It is of interest to assess the reductions in the solvent accessibilities of the different groups of atoms of the tRNA^{Phe} when it goes from a hypothetical, completely extended conformation to the folded crystal form. Such information may be useful for qualitatively understanding the kind of stability rendered by the solvent. Using the average values of the solvent accessibilities of the different atoms of each of the five types of nucleotide residues calculated for the crystal conformation, the average accessibility values for the specific atomic groups (namely, backbone-polar, backbone-nonpolar, base-polar, base-nonpolar, total-polar, total-nonpolar, and total) for each type of nucleotide residue was calculated. Similarly, the corresponding values were calculated for the extended conformation of the molecule using the accessibility values of the individual residues kept in the middle of a trinucleotide model. The ratios of the accessibility values of different atomic groups for the folded form to the corresponding quantities for the extended conformation of the tRNA^{Phe} were calculated, and the results are presented in table 1.

It is seen from table 1 that the reductions in the average accessibility values in the folded form for the normal nucleotide residues A, U, G, and C are in the same order for all the atomic groups except for the base-nonpolar atoms. This result is very interesting as it reveals that, in general, the adenosine residues are more buried while uridine residues are more exposed. The reductions found for the Y-nucleotide differ from the corresponding values found for the predominantly occurring nucleotide residues. For the backbone-polar and nonpolar groups of atoms, the reductions in the accessibility values when the molecule goes from the completely extended state to the crystalline form are, respectively, by factors of 1.4 and 1.1. This implies that the solvation effects of these atomic groups are not much altered due to folding process. However, for the base-polar and base-nonpolar atoms, for a corresponding phenomenon the reductions are, respectively, by factors of 4.7 and 6.9.

The significant reductions in the solvent accessibilities of the base atoms are mainly due to base pairing and other types of tertiary interactions. Because of these tertiary interchain interactions, most of the polar atoms of the bases are unable to associate with water. Consequently, this will also reduce the solvent accessibilities of the nonpolar atoms of the bases, as their ring systems are rigid and planar. However, the larger reduction in the case of nonpolar atoms of the bases compared to that of

Table 1. Reductions in the accessibilities of various classes of atoms in tRNA^{Phe} due to the transition from the extended state to the crystal form.

Nucleotide residue	Backbone		Bases		Total		
	Polar	Nonpolar	Polar	Nonpolar	Polar	Nonpolar	Total
A	0.725	0.877	0.179	0.108	0.488	0.302	0.429
G	0.709	0.952	0.210	0.141	0.448	0.400	0.437
U	0.676	0.847	0.261	0.232	0.500	0.422	0.478
C	0.685	0.971	0.215	0.140	0.481	0.403	0.461
Y	0.926	0.452	0.552	0.750	0.719	0.730	0.680
Ave.*	0.699	0.912	0.216	0.155	0.479	0.382	0.451

*for normal residues.

the polar atoms suggests the possibility of the existence of hydrophobic interactions in nucleic acids as earlier demonstrated by Crothers and Zimm (11) and by Ts'o (12).

When the total (backbone and bases) polar and nonpolar atomic groups are considered, the reduction factors for the folding process are found to be 2.1 and 2.6, respectively. These reduction factors are smaller than the values calculated for the polar and nonpolar groups of atoms in globular proteins where we found the reduction to be by a factor of 3 for both the kinds of atomic groups. When the total accessibilities of all the polar and nonpolar atomic groups were considered for the tRNA^{Phe} molecule, it is found that the polar and nonpolar groups of atoms occupy, respectively, 76% and 24% of the total accessible area available for the molecule. In globular proteins, it was found that the polar and nonpolar groups occupy more or less equal (50:50) solvent accessible areas.

The DNA Molecule

The DNA molecule was considered in the forms of five right-handed helical conformations (A, B, C, D, Lac B), two left-handed conformations (ZI, ZII), and one side-by-side (SBS) conformation. The atomic coordinates for these conformations were taken from literature. The unit of tetranucleoside triphosphate d-(GpCpGpC) in double helical form as shown in figure 2b was taken as a representative subunit since experimental data was available only for subunits with CG base pairs. The total accessible area of each of the eight helical states was taken to be the sum of the average accessible areas of one base pair and two nucleotide backbones. Hence it is assumed that this part represents one double helical nucleotide pair as marked in figure 2b. Also, we computed accessibilities not only with a probe of water molecule, but also with probes with radii ranging from 0 to 5 Å to investigate the problem in different dimensions.

The variations in the accessible areas of the eight helices when the size of the probe molecule is increased from a radius of 0 to 5 Å are shown in figure 7. We note from the plots of this figure that the DNA helices attain minimum surface areas not for a probe of radius 0.0 Å, but for probes with radii somewhere between 0.7 Å and 1.0 Å. This is a novel observation which enables us to measure the local surface roughness of the molecule when it adopts different helical forms: the probe radius at which the accessible surface area surpasses the van der Waals area is a direct measure of the roughness of the molecular structure. In such a measure, the relative roughness of the eight helices are given in table 2. From this table we note that the left-handed helices and the SBS helix are the smooth structures, and the D-helix is the most irregular structure; the A and B helices fall in one group and thus behave in a similar manner, though the relative exposures vary characteristically with the probe radius.

The contact areas of individual atoms bear a more physical meaning than the accessible areas since they represent the areas on which the probe molecule can have actual contact. In figure 8 we mark the available contact areas for probe molecules of radii equal to 0.0, 1.4, and 5 Å, the first case providing the van der Waals surface, the second corresponding to the action of a water molecule, and the last one providing information as to the existence of a bulk water structure around the molecule. We note from this figure that the two phosphate oxygens exhibit a remarkable difference in their contact areas for probe molecules of radii 1.4 and 5 Å. At 5 Å probe radius, the B-DNA alone exposes both of its phosphate

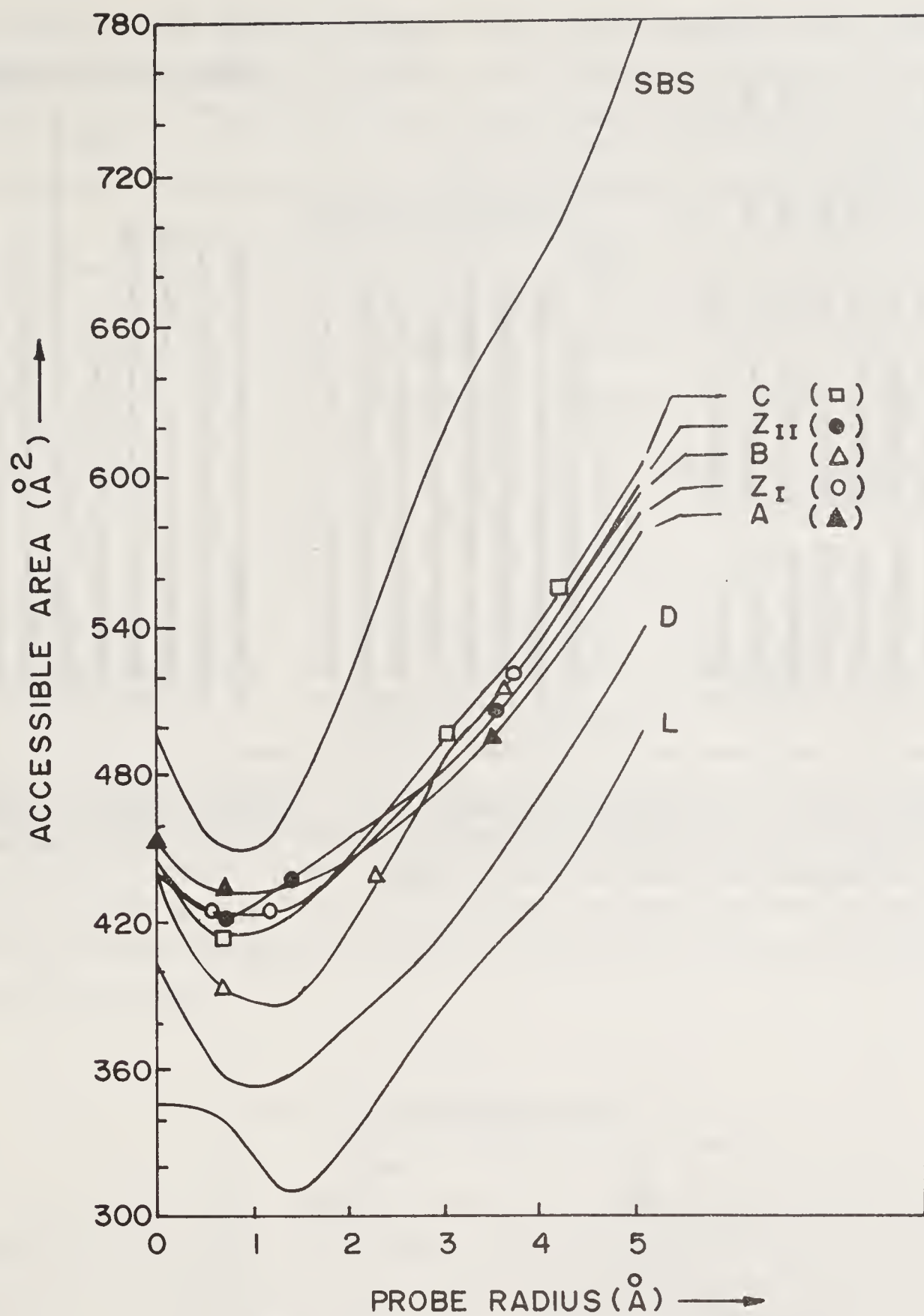


Figure 7. Total accessible areas of different types of helical DNA fragments for different probe radii.

Table 2. Relative roughness of the DNA helices.

Type of helix	Probe radius (Å) at which accessible area surpasses van der Waals area	Relative roughness
ZII	1.70	1.00
SBS	1.75	1.03
ZI	1.85	1.09
C	2.00	1.18
Lac B	2.25	1.32
B	2.30	1.35
A	2.35	1.38
D	2.75	1.62

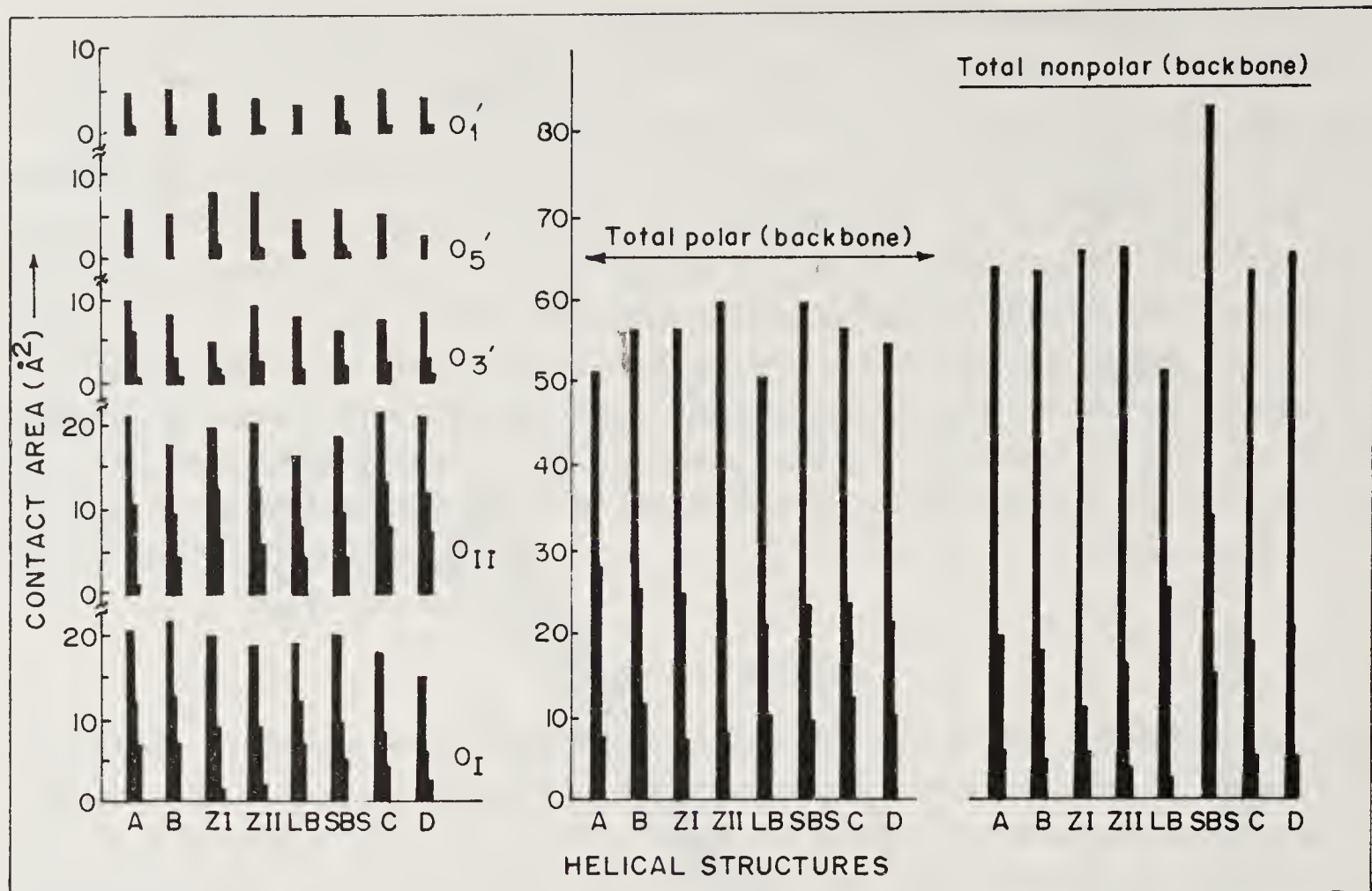


Figure 8. Contact areas of backbone atoms (OI, OII, ...) of DNA in eight different helical states; areas corresponding to probe radii 0.0, 1.4, and 5.0 Å are shown side by side against each helix from left to right.

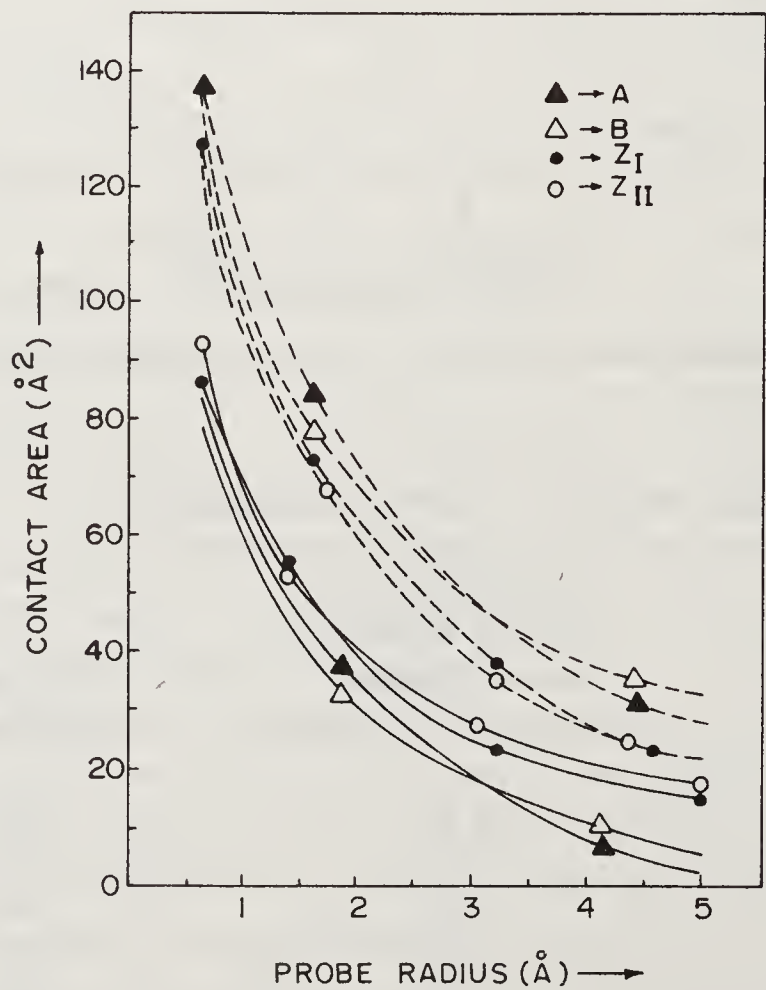


Figure 9. Total contact areas of different DNA fragments for various probe radii. Continuous lines correspond to base atoms and dashed lines to backbone atoms.

oxygens more to the probe molecule. This result indicates that there exists a secondary hydration shell of thickness 3–6 Å around this helical state and this bulk water structure may be the main reason for the preference of B-helix over the others. The A and Z helices expose greatly only one phosphate oxygen, keeping the other buried for probe molecules greater than the size of the water molecule. The decreased exposure of one of the phosphate oxygens in A and Z helices explain why these two helical families are less favored in high water solutions.

The variations of the contact areas of the backbone and base units with the increase in probe radius are shown in figure 9. Here we note that as the probe radius increases, the contact areas of both the backbone and base systems decrease rapidly. For a probe molecule of water size only 35 to 40% of van der Waals surface area of the backbone, and only 15 to 20% area of the base are amenable to exposure. Except for the Z helices, the base exposure is below 15% for the other helices. For a probe molecule of 5 Å radius, only 5 to 8% of the total surface area is accessible. Therefore, for a structured water, only a very small surface is accessible and that too mostly through the polar atoms. Interestingly, the A–B and ZI–ZII pairs undergo a transition in their backbone and base accessibilities, but at different probe radii. For a single water molecule, the ZII helix is more accessible, while for a structured water system the ZI helix is more accessible. The same result is noted for the A and B helices also. Another interesting feature is that the exposures of the backbone atoms are reduced by 25% when DNAs go from A and B helical states to Z helical states, whereas the base exposure is increased by a similar amount for a similar transition. This is because the bases in Z helices are pulled outward and they lie on the surface along with the backbone thus increasing their exposures and reducing the exposure of the backbone atoms.

Acknowledgement

I was a participant and a silent spectator for about five years of the events during the golden era of the Centre of Advanced Study in Physics, which Professor G. N. Ramachandran established, nourished and brought fame to by introducing the field of biopolymers to numerous youngsters at Madras University. One day after my seminar talk he called me to his room. I entered nervously, but was overtaken by his words, “keep it up”. That was the initiation day for my real research life. My later brief meetings with him shaped the course of my academic activities. This article is a small token of my gratitude to him.

I am thankful to Drs. P. Thiagarajan and A. Anukanth who collaborated in the work reported in this review article.

References

1. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379–400.
2. Richmond, T. J. & Richards, F. M. (1978) *J. Mol. Biol.* **119**, 537–555.
3. Chothia, C. (1974) *Nature (London)* **248**, 338–339.
4. Chothia, C. (1975) *Nature (London)* **254**, 304–308.
5. Ponnuswamy, P. K. & Thiagarajan, P. (1978) *Biopolymers* **17**, 2503–2518.

6. Thiyagarajan, P. & Ponnuswamy, P. K. (1979) *Biopolymers* **18**, 2233–2247.
7. Ponnuswamy, P. K. & Anukanth, A. (1984) *Curr. Trends in Life Sci.* **XII**, 31–37.
8. Jack, A. Ladner, J. E. & Klug, A. (1976) *J. Mol. Biol.* **108**, 619–649.
9. Stout, C. D., Mizuno, M., Rubin, J., Brennen, T., Rao, S. T. & Sundaralingam, M. (1976) *Nucleic Acids Res.* **3**, 1111–1124.
10. Bolton, P. H. & Kearns, D. R. (1978) *Biochim. Biophys. Acta* **517**, 329–337.
11. Crothers, D. M. & Zimm, B. E. (1964) *J. Mol. Biol.* **9**, 1–9.
12. Ts'o, P. O. P. (1970) in *Fine Structure of Proteins and Nucleic Acids*, eds. Fasman, G. D. & Timasheff, S. N. (Dekker, New York), pp. 166–168.

DNA Bending: A Natural Consequence of Base Sequence Dependent Variability

Manju Bansal, Dhananjay Bhattacharyya and
Debasisa Mohanty

Molecular Biophysics Unit, Indian Institute of Science,
Bangalore 560 012, India

Introduction

It is difficult to believe that, a short while ago, DNA molecules were thought of as being rather rigid, uniform and rod-like, which undergo bending or distortion only under conditions of stress. While the B-form was supposed to be the predominant structure *in vivo* and quite flexible, the A-form was supposed to be very rigid and its biological significance was also not so well determined. Even the well characterized structural transitions between $A \rightleftharpoons B$, or $B \rightleftharpoons C$, D etc were supposed to occur between intrinsically straight molecules with minor variations in their diameter and elongation. From purely stereochemical considerations of polynucleotide chains, such a view is completely untenable. On the basis of theoretical conformation studies of mononucleotides and basepaired dinucleotides it was postulated more than ten years ago from our laboratory that the DNA structure is intrinsically flexible and more than one model can be generated to fit the X-ray data available from fiber diffraction studies of DNA (1–4). As a natural corollary to this hypothesis, it was suggested that even the sense of the DNA helix need not necessarily be always right-handed, but that left-handed helices could also be built and are in reasonably good agreement with the X-ray diffraction as well as infra-red data from DNA fibers (3, 4). However until the elucidation of Z-DNA structure for d(CGCGCG), the idea of a left-handed structure was considered blasphemous, in spite of the fact that Watson–Crick as well as Wilkins and co-workers in their original papers had not ruled out the possibility of such a structure for B-DNA (5–7).

In the last few years the concept of sequence dependent variability in DNA structure as well as intrinsic flexibility for each basepair doublet has become the accepted norm, rather than an exception seen only under unusual environmental or functional conditions. While hydrodynamic and gel-electrophoretic studies give an indirect indication that the DNA molecule is not always a straight entity (8–13), the X-ray diffraction analysis of oligonucleotides provides direct evidence of the range of conformational flexibility that is possible in DNA structure (14–16). However, even the crystal structures provide only a static snapshot of one of the possible conformational states of the molecule and is likely to be influenced by intermolecular and packing constraints imposed by the crystalline lattice (16–17). Therefore, detailed

theoretical studies of basepair stacking preferences are also essential to understand the full range of conformations accessible to the DNA molecule (18–22). Both intrinsic and induced bending of DNA molecules is likely to be facilitated by sequences which can readily undergo fluctuations from the ideal straight structure. An intrinsic, static bend will occur at sequences which have a well defined energy minima for doublet geometries that are considerably different from other flanking sequences, while sequences with wide shallow minima will allow structural fluctuations leading to transient bends, as well as being preferred sites for ligand induced bends. In this paper some of our recent work on stacking energy calculations, as well as analysis of base sequence effects in DNA oligonucleotide crystal structures is described. Its implication and application to DNA bending in solution is discussed with particular reference to the gel migration data of some naturally occurring DNA sequences.

Method

In order to understand the base sequence dependent features of nucleic acid molecules, first a complete description of all possible doublet steps consisting of two basepairs is necessary. Such a description requires a set of intra- and inter-basepair doublet parameters, each consisting of three rotational and three translational parameters (23). A generally acceptable parameter set along with its nomenclature was evolved at the EMBO workshop on “DNA Bending and Curvature” held at Cambridge, UK in 1988 (24, 25). Of the six intra-basepair parameters, the inter-base translations show relatively small variations and are not discussed here. The three inter-base rotations are termed as propeller, buckle and opening angle and significantly affect the inter-basepair geometries (26).

The set of local inter-basepair parameters consists of three rotations: tilt (τ), roll (ρ), twist (ω_1) and three translations: shift (Dx), slide (Dy) and rise (Dz) (24). We have recently developed a self consistent algorithm (called ‘NUPARM’) for the analysis and generation of all types of non-uniform, including smoothly curved or kinked nucleic acid fragments (25, 26). In our description, the local parameters for a doublet step are defined with respect to a mean plane lying between the two constituent basepairs and are related to another set of six parameters which are defined with respect to a local helix. Both sets of local parameters are:

- (i) independent of the orientations of the neighbouring doublets,
- (ii) independent of the direction in which the doublet is taken, except for a sign change in tilt and shift or the related parameters tip and dy (i.e. same values are obtained when the step is considered from residue $i \rightarrow i + 1$ of either strand). Hence the parameters can be compared for nucleic acid fragments of varying length, sequence and tertiary structural arrangement. The values for the local helical parameters are generally similar to the conventional parameters, displacement (dx), inclination (η), helical twist (ω_h) and rise (dz), which are used to describe uniform double helical structures (27, 28). The differences in the stacking geometry and basepair overlap seen in the fiber models of A and B forms of DNA are discussed along with oligonucleotide crystal structures. Detailed theoretical stacking energy calculations have been carried out for all ten doublet sequences, using the all atom AMBER (Assisted Model Building with Energy Refinement) force field (29, 30) and varying the doublet geometry using the local step parameters and compared with the results

of crystal structure analysis. The mean values obtained for the local doublet parameters for the ten doublet sequences have been used to generate fragments of naturally occurring DNA and various geometric parameters calculated, so as to characterize the molecule as being straight or curved.

Results and Discussion

Analysis of A and B DNA Models Obtained from Fiber Diffraction Data

In the well known fiber models of DNA, the basepair doublet parameters, like all other variables have to necessarily repeat periodically to give a uniform helical model. Since the local doublet parameters are related to the local helical parameters, both the sets of parameters have constant values for all the steps of a fiber model with mononucleotide repeating unit. The intra-basepair propeller twist along with the local step and local helical parameters, for several A and B-DNA fiber models (2, 3, 31–33) are listed in table 1 a and b. Since the fiber models have been generated assuming that the two strands are related by a pseudo-dyad, coinciding with the x-axis, the local step parameters tilt (τ) and shift (Dx), the related local helical parameters tip (Θ) and dy and the intra-basepair parameter buckle are nearly zero in all these models.

The two representative models listed for the right handed A form of DNA have basepair inclination of 19° and 23° respectively while the propeller twist changes from

Table 1. Local doublet and helical parameters of the DNA models proposed on the basis of X-ray fiber data.

(a) Basepair propeller and local doublet parameters.

	Propeller (π)	Tilt (τ)	Roll (ρ)	Twist (ω_1)	Shift (Dx)	Slide (Dy)	Rise (Dz)
A-DNA (31)	11.74	−0.07	11.09	30.84	0.06	−1.50	3.27
A-DNA (32)	−10.51	−0.11	13.13	30.04	0.06	−1.22	3.34
B-DNA (31)	4.24	−0.06	−3.85	35.80	−0.06	−0.23	3.38
B-DNA (33)	−13.33	0.01	1.32	35.98	0.05	0.60	3.36
B-DNA (3)	3.04	0.08	−2.12	35.94	0.05	0.20	3.42
B-DNA (2)	−12.64	0.04	−5.47	35.60	0.04	−0.68	3.33

(b) Helical parameters.

	Inclination	Tip	Twist	Displacement		Rise
	(η)	(Θ)	(ω_h)	dx	dy	(dz)
A-DNA (31)	19.31	0.13	32.73	−4.74	−0.12	2.57
A-DNA (32)	23.08	0.19	32.73	−4.74	−0.13	2.57
B-DNA (31)	−5.94	0.06	36.00	0.23	0.09	3.38
B-DNA (33)	2.03	−0.02	36.00	0.76	−0.09	3.38
B-DNA (3)	−3.26	−0.12	36.00	0.65	−0.07	3.40
B-DNA (2)	−8.45	−0.06	36.00	−0.26	−0.07	3.40

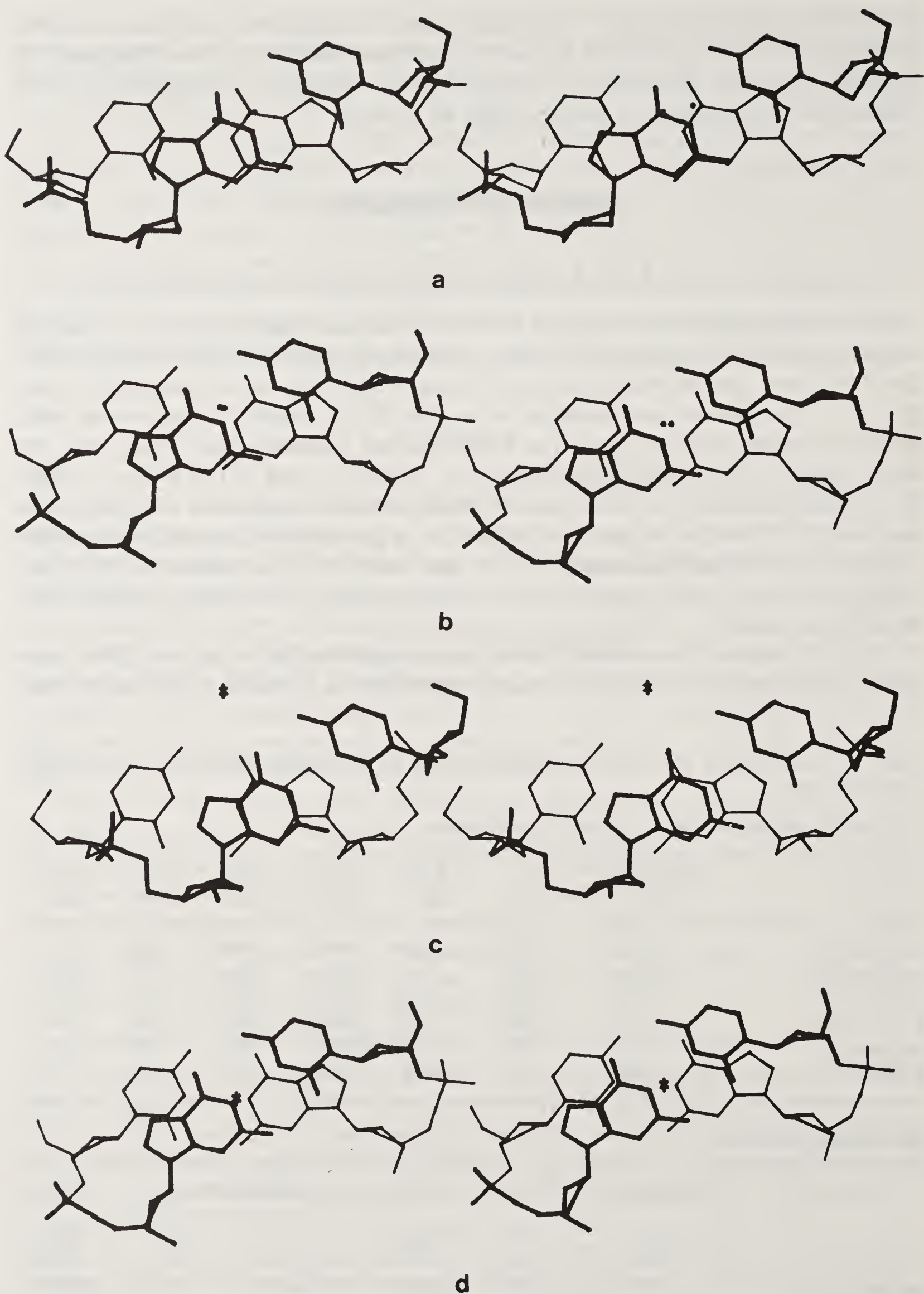


Figure 1. Stereo projection down the mean z -axis (Z_m) of the basepaired dinucleotide d(CG).d(CG) for the (a) A-DNA and (b) B-DNA fiber models. The symbol ‘.’ represents the basepair centers in each case. (c) and (d) show the same doublet structures projected down the helix axis (Z_h) corresponding to the A- and B-forms respectively. ‘*’ indicates position of the helix origin.

12° to -11° . Both the A-DNA models, which are in addition characterized in the helical framework by a large negative x -displacement ($\approx -4.7\text{\AA}$) of the basepairs, are described in the mean local axial system by large negative values for slide ($\approx -1.5\text{\AA}$ and -1.2\AA) and positive roll (11° and 13°). It is interesting to note that such a large difference in the propeller twist and quite significant variation in the roll and slide values for the two models (table 1), can still give rise to A-DNA models which are supposed to fit the A-DNA X-ray fiber data equally well. This is in agreement with our earlier report that even the most crystalline X-ray fiber data (as in A-DNA) determines precisely the helical parameters (rise and twist), and gross parameters such as phosphate radius and chain separation across the grooves, but can be fitted to structures with different local conformations (34).

The various right handed B-DNA models listed in table 1, as well as its minor variant for C-DNA, have small basepair inclination and correspondingly small roll angles. However, while the x -displacement is generally positive and varies from 0.23\AA to 0.76\AA for these three structures, the slide parameter fluctuates from small negative to small positive values. The B-DNA model with C3'-endo sugar pucker however has small negative values for displacement as well as slide, as in A-DNA, but the roll angle and basepair inclination are negative. In general the variation in the magnitude of the local step parameters describing the relative translations and rotations between the two basepairs is small, so as to maintain the basepair planes nearly parallel to each other and preserve favourable base-stacking interactions. However, when the basepairs are described with respect to the imaginary doublet helix axis and the twist has a small value, minor variations in roll and slide values translate into large values of inclination and x -displacement.

It is also interesting to see that the projections down the mean z -axis (Z_m) (figure 1a and b) for the A and B structures do not appear to be as dramatically different as the projections down the helix axis (Z_h) (figure 1c and d). In the case of B-DNA models, the basepairs in a doublet step are nearly parallel to each other, leading to the Z_m and Z_h axes being also aligned in almost parallel directions and the projections down Z_m and Z_h are quite similar (figures 1b and 1d). However, for the A-DNA structure the inclination of the basepairs combined with their relative twist and sliding movement leads to the neighbouring basepairs becoming non-parallel and basepair centers are moved about 4.7\AA away from the helix origin. When viewing down the Z_m -axis, the amount of intrastrand overlap appears to be greater while the interstrand overlap appears to be much less when compared to the usual projection (figure 1c) down the local helix axis.

Base Sequence Effects in B-DNA Oligonucleotides

The local doublet parameters in various B-DNA dodecamer structures with the sequence d(CGCGAATTCGCG) but under different conditions and with various bound ligands, have been analysed in detail elsewhere (26). It is concluded that the presence of bound drug or water molecules has only a minor effect on the local geometry of various basepair doublets while the base sequence and the position of the doublet are more crucial. In order to study the effect of different base sequences in similar crystalline environment, the local parameters for four dodecamers of type d(CGCxxxxxxGCG) with the sequences (GAATTC), (AAATTT), (GATATC) and (ATATAT) in the middle six positions (35–38) have been examined in detail here. As

in the earlier analysis (26) the intra-basepair propeller twist has slightly larger values for the A:T basepairs compared to the G:C basepairs while opening angle and buckle are slightly smaller. The shift and tilt parameters have small values throughout the molecules, as expected for pseudodyad related double helical structures, while roll and slide show quite large fluctuations, particularly at the terminal basepair steps (figure 2a, b). It is interesting to note that the sequence effect between AA, AT and TA steps is only marginal for both these parameters and slide in particular retains its preference for negative values at these central steps, even when the drug molecules are bound in the minor groove. Even the alternating AT and TA steps show no marked preference for alternating values of roll, unlike the GC and CG steps at the ends. On the other hand rise and particularly twist have alternating values for AT and TA, as well as GC and CG sequences (figures 2c and d), but with opposite trends.

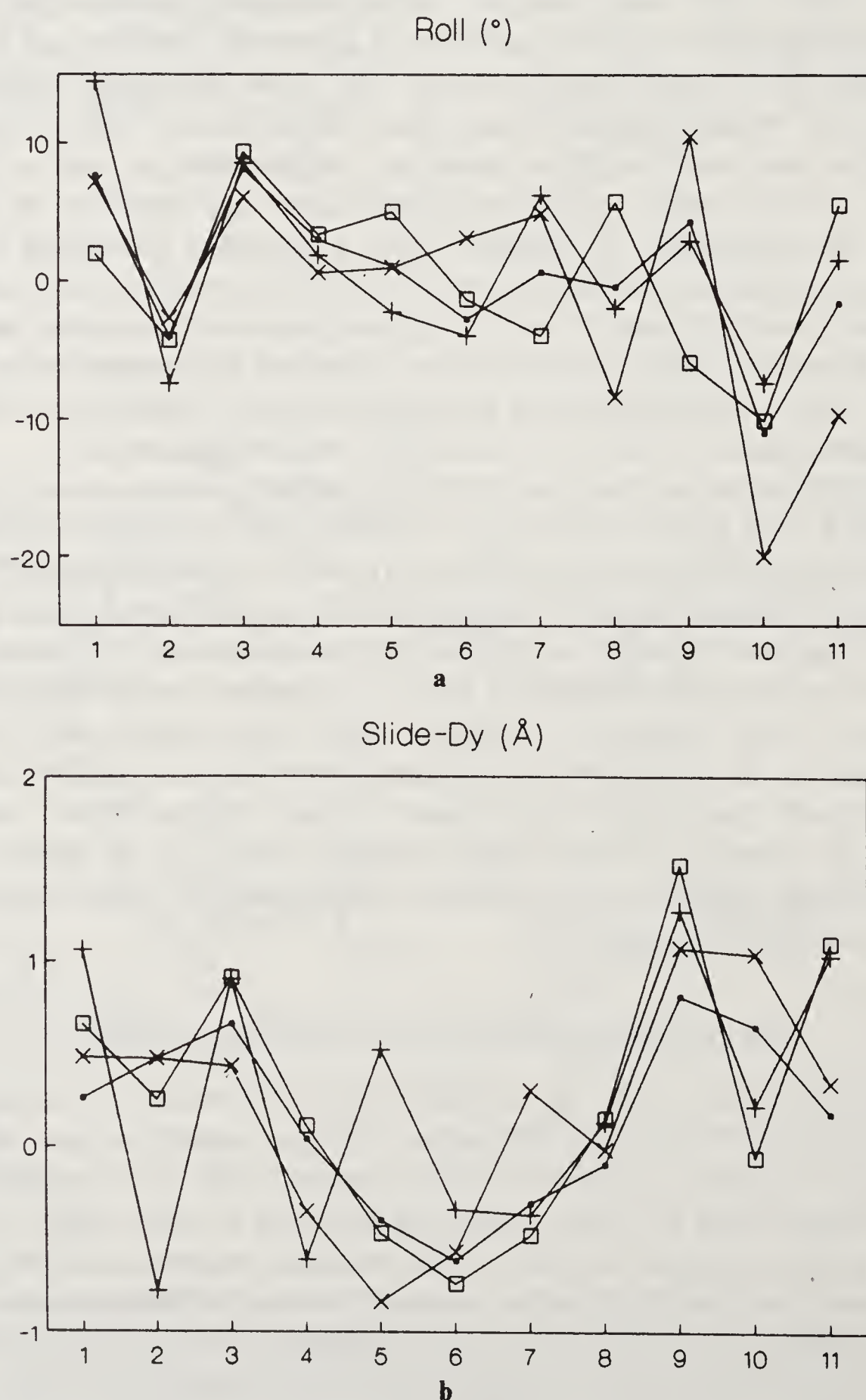


Figure 2. (a) and (b) (see facing page for caption).

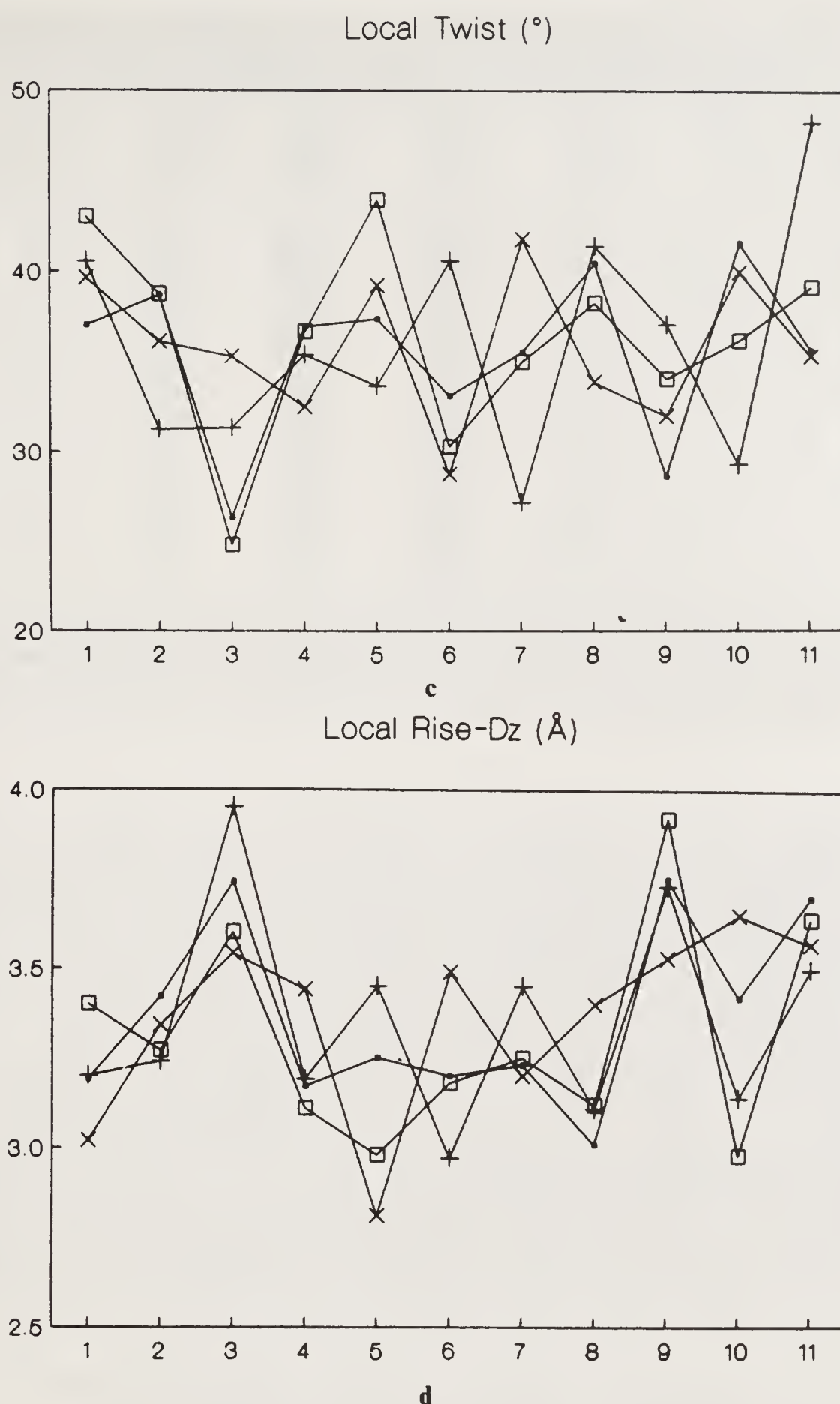


Figure 2. Variation of the doublet parameters (a) roll, (b) slide, (c) local twist and (d) local rise for four dodecamer structures with the sequences d(CGCGAATTCGCG), native structure, d(CGCGATATCGCG) with Netropsin, d(CGCAAATTTGCG) with Distamycin and d(CGCATATATGCG) are shown, represented by the symbols ‘.’, ‘+’, ‘□’ and ‘×’ respectively.

The variation in rise and twist is at least partly due to the opposite buckling of the constituent basepairs in the doublet and if a correction is applied for the buckling effect, the local rise has a nearly constant value for these sequences (26). The mean values and the standard deviations of the local doublet parameters, have been calculated for all the B-DNA structures, whose coordinates are available in July 1990 release of Brookhaven Protein Data Bank (39) and are listed in table 2. The most striking feature in this table is the relatively large twist at GA, GC and TA steps,

Table 2. The mean values (for the ten doublet sequences) of the six local parameters, tilt (τ), roll (ρ), twist (ω_1), shift (Dx), slide (Dy) and rise (Dz), of B-DNA are listed here.

	No. of steps	Tilt (τ)	Roll (ρ)	Twist (ω_1)	Shift (Dx)	Slide (Dy)	Rise (Dz)	Corrected ⁺ Rise
GG	3	0.82	6.05	30.78	− 0.38	0.62	3.47	3.39
AA	25	0.64	2.96	36.46	− 0.06	− 0.21	3.17	3.27
AG	1	− 0.46	6.94	23.71	0.79	0.94	3.65	3.41
GA	22	− 0.31	0.01	38.41	0.02	− 0.07	3.12	3.27
GC	26	− 0.31	− 8.71	38.73	− 0.26	0.56	3.34	3.47
AT	16	0.13	− 0.75	31.56	0.05	− 0.45	3.27	3.25
AC	1	1.59	− 2.29	29.40	− 1.19	− 0.05	3.62	3.35
CG	48	− 0.29	3.68	33.99	0.07	0.63	3.64	3.47
TA	4	− 0.03	2.43	39.15	− 0.01	− 0.26	3.16	3.33
CA*	3	− 1.04	8.64	30.69	0.12	0.87	3.56	3.41
	3	1.13	− 7.80	44.83	0.05	2.24	3.56	3.38

⁺ The local rise values obtained after correcting for the buckling in the basepairs constituting the doublet.
* For the CA step, two mean values are listed, since two widely different types of structures are observed for this dinucleotide sequence, one with the normal backbone conformation and the other with B_{II} backbone conformation. The second line gives the parameters corresponding to the steps with B_{II} conformation.

with the GC step having a large negative roll also. A slightly smaller twist is observed for GG, AT and CG steps. The mean local rise Dz which shows considerable variation for the ten doublets, becomes nearly constant if a correction is applied for the effect of basepair buckling (26). The AA, AT, TA and GA sequences show slightly smaller rise values compared to the other sequences (as seen in last column of table 2). The CA/TG sequence shows two types of geometries, one with positive roll, small positive slide and reduced twist, and another with negative roll, large positive slide and very large twist. Some of the other sequences also show considerable deviation from B-DNA fiber geometry (viz. AG and AC steps), but there are very few steps with these sequences in the present data base. Some marked differences are also seen for the same sequences when they occur in decamer and dodecamer crystals, presumably due to the different intermolecular interactions in the crystal lattice.

Base Sequence Effects in A-DNA Oligonucleotides

Only seven doublet sequences are present in the available oligonucleotide data base for A-type DNA structures (26, 39–41). Since similar base sequence effects are seen in double helical regions of crystalline tRNA^{Phe} and tRNA^{Asp} molecules, we have calculated the mean values of the local doublet parameters for the A-DNA oligonucleotide structures as well as tRNA structures taken together and these are given in table 3. It is striking to see that most sequences show a smaller roll and slide compared to the models proposed from X-ray fiber data (table 1a). Only the pyrimidine–purine sequences CA, TA and CG, show large roll and slide values accompanied by small twist. As expected from the theoretical relations given by us (25, 26), this combination of local step parameters leads to very small values of helical rise in addition to the large basepair displacement and inclination, characteristic of A-DNA fiber models (table 1b). On the other hand, all the three purine–pyrimidine

Table 3. Mean doublet parameters for all the doublet sequences in A-form structures, i.e. A-DNA and tRNA structures available in PDB file (39) and refs. (40, 41).

	No. of steps	Tilt (τ)	Roll (ρ)	Twist (ω_1)	Shift (Dx)	Slide (Dy)	Rise (Dz)
GG	31	-0.72	6.11	31.86	-0.07	-1.74	3.32
UU	7	-0.39	13.99	32.70	-0.22	-1.23	3.27
AG	15	0.82	5.84	32.15	-0.20	-1.68	3.24
GA	16	3.09	6.55	29.57	0.82	-1.36	3.31
GC	16	-2.13	5.09	32.49	-0.32	-1.37	3.28
AU/T	8	0.31	5.97	31.42	0.16	-1.33	3.34
AC	13	-0.48	4.89	32.26	0.08	-1.20	3.47
CG	12	2.15	12.34	31.00	0.02	-1.59	3.10
	2*	0.00	4.49	22.06	0.00	-2.19	3.64
U/TA	6	0.09	12.51	27.98	-0.13	-1.41	3.31
CA	16	0.94	8.19	30.17	0.23	-1.65	3.18

*These steps are the sites of a molecular dyad axis in the oligonucleotides and have *trans, trans* conformation for the backbone torsion angles α, γ instead of the usual *gauche*⁻, *gauche*⁺ conformation.

sequences AC, AT and GC are characterised by relatively smaller roll (5° to 6°) and slide (-1.2Å to -1.4Å) values with normal twist angles ($\approx 32^\circ$). Consequently the local helical rise is almost like B-DNA while the basepair inclination is also quite small. The homopurine steps GG and AG have the combination of small roll, large slide and normal twist, leading to basepair displacement and inclination values which are close to the A-DNA fiber models but with a helical rise of about 2.9Å. The AA sequences are only available from the tRNA crystal structures and hence correspond to A:U basepairs. It shows the unique combination of large positive roll, reduced slide and normal twist.

Thus the doublet geometries in A form structures with C3'-endo sugar pucker show a slightly smaller range, but have much more stronger sequence dependence than in B form. However neither in A nor B form are any of the local parameters found to vary in a correlated manner (26). No strong correlation is found between the local parameters and the backbone torsion angles within either of the two types of structures, though a change in the backbone conformation of both strands in the doublet step, such as B_I to B_{II} in B-DNA or *gauche*⁻, *gauche*⁺ to *trans, trans* conformations of torsion angles α, γ in A-form, leads to noticeable differences in the basepair doublet parameters.

Stacking Energy of Basepair Doublets

The stacking energy of basepair doublet sequences has been calculated earlier by several other groups, for the B-DNA or A-DNA type structures (19–22). However, generally the helical parameters are considered as variables and only minor fluctuations from the fiber models have been examined. We have calculated the stacking energy of all ten doublet steps by varying the four local doublet parameters roll, twist, slide and rise. The intra-basepair propeller twist was also varied. It was found to improve the intrachain nonbonded energy component (42) when it has a negative value as seen in most crystal structures.

Since the local z -rise remains almost constant (at a value of 3.3 \AA) in all DNA structures we restricted our calculations to rise values varying from 3.2 to 3.4 \AA . Local twist was not found to have any significant effect on the nonbonded stacking energy in the whole region from 20° to 40° . It was therefore fixed at 30° (to simulate A-DNA structures) and 36° (to simulate B-DNA structures). The stacking energy has a slightly more favourable value (by about 0.5 kcal/mole) for the smaller twist value, for all sequences. We then examined the stacking energy as a function of roll *vs* slide variation, over the range of $\pm 25^\circ$ in roll and $\pm 2.5 \text{ \AA}$ in slide. Representative stacking energy plots for six doublet sequences GG, AA, GC, AT, CG and TA are shown in figure 3, for 0° propeller, 3.3 \AA rise and 36° twist. It is seen that for most sequences the energy minima occurs close to 0° of roll and 0 \AA of slide, with the 2 kcal total energy contour encompassing roll angle fluctuations of $\pm 8^\circ$ for zero slide and allowing slide values to vary between -1.5 to $+1.5 \text{ \AA}$ for 0° roll angle. However for some of the sequences a correlated variation in roll and slide allows much larger fluctuations, for example for AA sequence, negative roll with negative slide as well as the combination of positive roll with positive slide allow roll angle to have values of -15° and $+15^\circ$ respectively, while positive roll with negative slide values, characteristic of A-DNA, is less favourable (figure 3).

For all the three pyrimidine–purine sequences larger values of positive roll are allowed compared to negative values, for a wide range of slide, both from nonbonded as well as electrostatic energy calculations. The opposite trend is seen for the three purine–pyrimidine sequences with slightly larger negative roll values being allowed for most values of slide (figure 3). This is in agreement with the trends seen in B-form crystal structures, particularly for CG and GC sequences (table 2). A rather surprising result of the stacking energy calculation was the finding that the basepair orientation in A-DNA fiber models is invariably $2\text{--}3 \text{ kcal/mole}$ higher in energy compared to the absolute energy minima, both for nonbonded energy alone, as well as when electrostatic energy is included. The difference reduces by about 1 kcal for GG and all three pyrimidine–purine sequences when a propeller twist of -10° is applied (figures 3 and 4). Since the A-DNA fiber models correspond to roll angle of about 10° and slide of -1.5 \AA , it is obvious from the earlier discussion, that the energy for the purine–pyrimidine sequences with such a large positive roll is substantially higher from their global minima, when compared to the pyrimidine–purine sequences. Hence, it provides a theoretical rationale for the base sequence effects seen in A-DNA as well as tRNA double helices (table 3), and the observed deviations in the crystal structures, for all the doublet sequences, serve to improve the total stacking energy of basepairs in A-DNA helices. This is clearly seen in figure 4a for the octamer d(GGTATACC) structure in A form, wherein the GG, GT and AT steps improve in energy by almost 2 kcal over their stacking energy in the fiber geometry.

In the case of B-DNA the results are not so clearcut, since any deviation from near 0° roll and 0 \AA slide makes the stacking energy less favourable. However, in this case also the mean geometries for all doublet sequences (as given in table 2) are found to be within 1 kcal/mole contour, though some individual steps deviate by upto 2 kcal from the energetically optimum geometry (see figure 4b). While some of the deviations can be attributed to intermolecular interactions, for example the steps 10 and 11 in the B-DNA dodecamers, some others cannot be explained by any obvious source of perturbation. Several of these steps have the unusual B_{II} conformation corresponding to *gauche*⁻, *trans* orientation for torsion angles

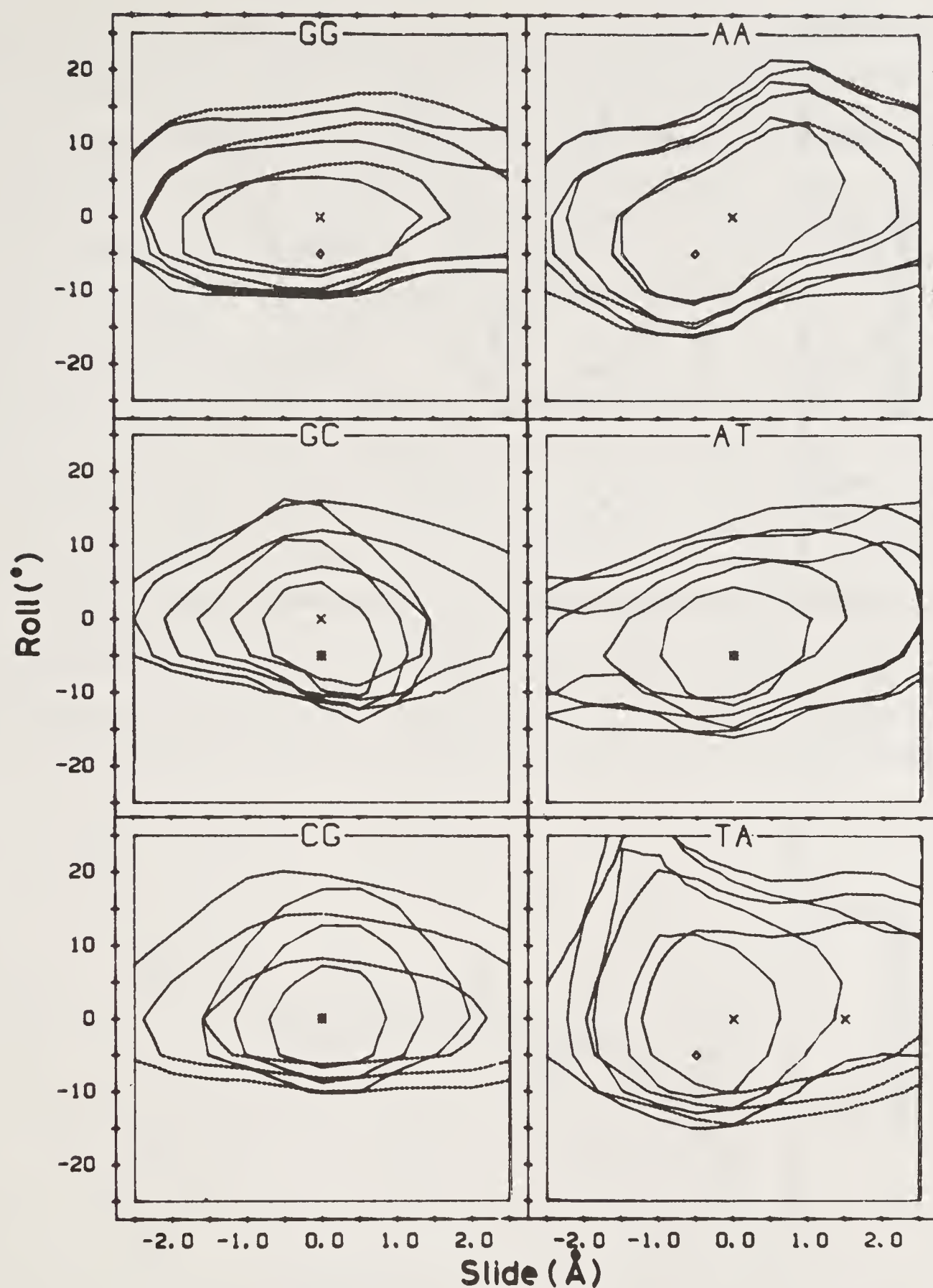


Figure 3. The stacking energy contours for six doublet sequences are shown as a function of roll and slide variation, with fixed rise (3.3 Å), twist (36°) and 0° propeller. The line types '—' and '...' represent the contours corresponding to total stacking energy (nonbonded + electrostatic) and nonbonded energy alone, with the '□' and 'x' symbols representing the positions of minimum energy. Three contours, 1.0, 2.0 and 3.0 kcal/mole above the respective minima have been shown in each case.

$\epsilon(\text{C3}' - \text{O3}')$, $\zeta(\text{O3}' - \text{P})$ which leads to favourable sugar-phosphate interaction at the 3'-end (43) and this may be compensating for the loss of base stacking energy at these steps.

Hence in general the base sequence effects seen in crystal structures are in agreement with the trends predicted by theoretical stacking energy calculations. We have therefore used the mean values obtained from crystal structure analysis to study the trajectory of some naturally occurring DNA fragments.

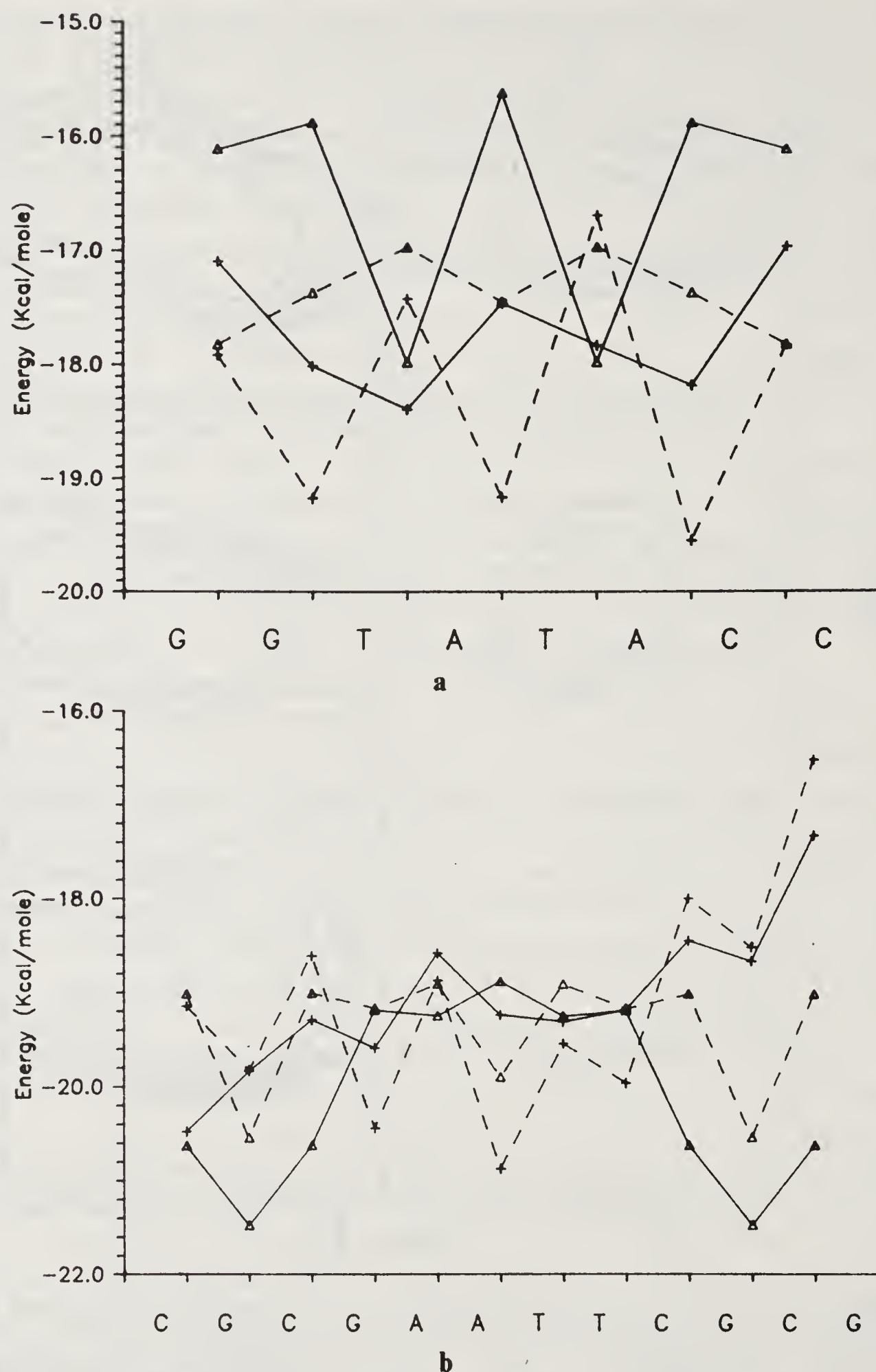


Figure 4. The total '—' and nonbonded '- - -' energy for each doublet step is shown for representative (a) A-DNA and (b) B-DNA structures. The values corresponding to the crystal geometry are indicated by the symbol '+' while the fiber model geometry is represented by 'Δ'.

Base Sequence and DNA Bending

Basepair doublets were generated using the mean values of the local doublet parameters, obtained for the available crystal structure data in both A and B-form geometry, and listed in tables 2 and 3. The sequences chosen were those corresponding to the restriction fragments of kinetoplast DNA from *Crithidia fasciculata* (I, II and

III) and *Leishmania tarentolea* (IV) as well as the 124 basepair sequence at the origin of replication of bacteriophage λ (V) (44–46). The trajectory of the basepair centers in both A and B forms for all four sequences is shown in figure 5. It is interesting to see that the same fragments (II and IV) appear to be bent in both forms (figure 5a) while fragments I and III are relatively straight (figure 5b). These trends are in agreement with the reported bending of these fragments from gel migration studies, as seen from the values of relative mobility (R_1) listed in table 4. The ratio of the end-to-end distance to the path length has significantly smaller values for fragments II and IV, as also the ratio of the largest to smallest moments of inertia of the generated molecules. Both these parameters indicate that these fragments are considerably compressed when compared to the other three sequences examined. The fragments II and IV are characterized by long runs of adenines and it was initially proposed that the large rolling and tilting of AA steps causes them to bend (47). However, Crothers and coworkers from gel migration studies proposed that it is the large tilt at the junction between the A rich regions with random sequence which causes the bend (13). It is interesting to see that two widely different sets of local doublet parameters, with very little tilt variation, predict the same sequences to be intrinsically curved. A third set proposed by Olson and coworkers from theoretical considerations also predict similar trends (20). All these parameters have a range of variation in roll angles while tilt is invariably small. However our parameters, from

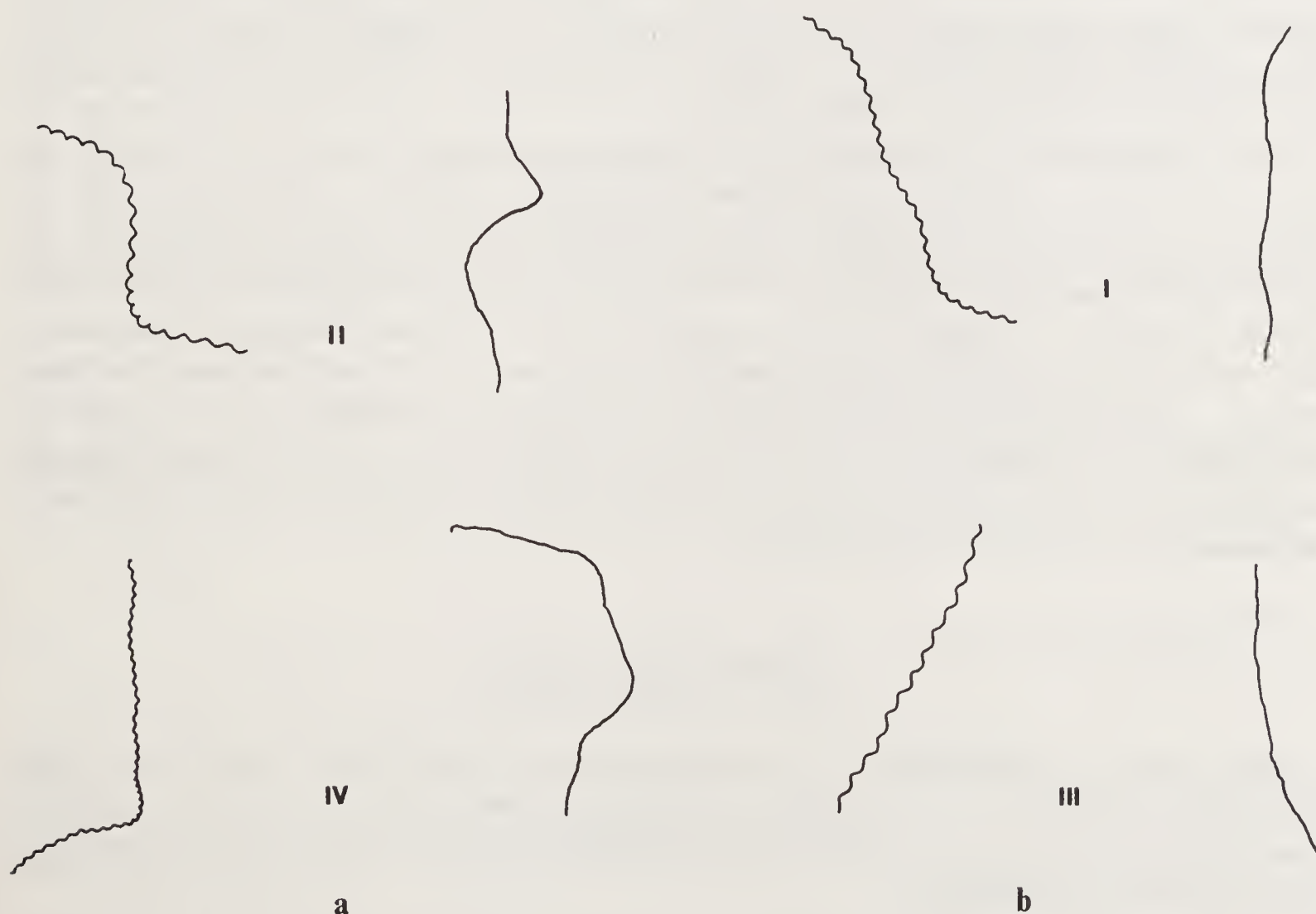


Figure 5. The basepair centers of the theoretically generated DNA molecules with the sequences corresponding to the restriction fragments (I–IV) from kinetoplast DNA are shown. The models with the local doublet parameters for A-form (table 3) are shown on the left and those with B-form geometry are on the right. The fragments II and IV are curved in both cases (a) while fragments I and III are relatively straight (b) in agreement with the data given in table 4.

Table 4. The experimentally determined values of the relative retardation in gel mobility of five restriction fragments of DNA compared with the geometric dimensions of the theoretically generated models, using B-form and A-form doublet parameters, as listed in tables 2 and 3 respectively.

Fragment number	No. of basepair	Expt. R_1^s	B-form parameters		A-form parameters	
			$(d/l_{\max})^+$	$(I_{\max}/I_{\min})^*$	(d/l_{\max})	(I_{\max}/I_{\min})
I	274	1.09	0.95	109.83	0.84	24.05
II	211	3.18–5.21	0.81	15.13	0.64	5.28
III	150	1.00	0.97	58.62	0.97	111.08
IV	414	3.78	0.74	7.30	0.70	7.63
V	124	< 1.17	0.96	48.43	0.81	13.61

^s R_1 is the reported experimental ratio of apparent to actual molecular size from gel migration studies (44–46).
[†]This is the ratio between the end-to-end distance (d) and the path length (l_{\max}). $l_{\max} = \text{mean helical rise times the number of doublet steps}$, where the mean helical rise from crystal structure data has values of 3.32 Å and 2.90 Å for B and A forms respectively.
^{*} I_{\max} and I_{\min} are the largest and smallest values of the principal moments of inertia of the generated structures.

crystal structure analysis also have a twist variation while the theoretical values were derived using a fixed value of twist. The values of roll angles calculated by Olson’s group (20) are invariably positive (as in A-DNA) and hence the amount of relative compression predicted by them is closer to the values calculated by us using the A-form parameters, even though there are significant differences in the values for the various doublets. The fact that all three parameters lead to similar predictions for DNA sequences bending indicates that the actual values of the roll angles are not critical, but the relative changes in roll between the different doublet steps determines which sequences will bend. We also calculated the structural comparison parameters for random sequences of the same lengths as fragments I–V and find that the values of d/l_{\max} are generally greater than 0.9, indicating that the small values obtained for the kinetoplast fragments II and IV are true reflections of their unusual basepair sequences. A more detailed analysis including data for synthetic polynucleotides is underway to confirm this hypothesis.

Acknowledgements

This work was supported by financial assistance from CSIR, India. We are also grateful to Drs. Dickerson, Heinmann, Kennard and Wang for providing unpublished coordinate data and to the Distributed Information Center, I.I.Sc., India for providing the Brookhaven PDB data.

References

1. Sasisekharan, V. & Pattabiraman, N. (1978) *Nature (London)* **275**, 159–162.
2. Gupta, G., Bansal, M. & Sasisekharan, V. (1980) *Int. J. Biol. Macromol.* **2**, 368–380.

3. Sasisekharan, V., Bansal, M. & Gupta, G. (1983) in *Nucleic Acids: The Vectors of Life*, eds. Pullman, B. & Jortner, J. (Reidel Publishing Co., Dordrecht), pp. 101–111.
4. Bansal, M. & Sasisekharan, V. (1986) in *Theoretical Chemistry of Biological Systems*, ed. Naray Szabo, G. (Elsevier, Amsterdam, New York), pp. 127–218.
5. Watson, J. D. & Crick, F. H. C. (1953) *Nature (London)* **171**, 737–738.
6. Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Cooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. (1960) *J. Mol. Biol.* **2**, 38–64.
7. Fuller, W., Wilkins, M. H. F., Wilson, H. R., Hamilton, L. D. & Arnott, S. (1965) *J. Mol. Biol.* **12**, 60–80.
8. Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7664–7668.
9. Hagerman, P. J. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81**, 4632–4636.
10. Wu, H.-M. & Crothers, D. M. (1984) *Nature (London)* **308**, 509–513.
11. Hagerman, P. J. (1986) *Nature (London)* **321**, 449–450.
12. Diekmann, S. (1986) *FEBS Lett.* **195**, 53–56.
13. Koo, H.-S. & Crothers, D. M. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1763–1767.
14. Dickerson, R. E., Kopka, M. L. & Pjura, P. (1985) in *Biological Macromolecules and Assemblies*, eds. Jurnak, A. & McPherson, A., Vol. 2, pp. 37–126.
15. Dickerson, R. E. (1990) in *Structure and Methods: DNA and RNA*, eds. Sarma, R. H. & Sarma, M. H. (Adenine Press) Vol. 3, pp. 1–38.
16. Shakked, Z., Guerstein-Guzikevich, G., Zaytzev, A., Eisenstein, M., Frolov, F. & Rabinovich, D. (1990) in *Structure and Methods: DNA and RNA*, eds. Sarma, R. H. & Sarma, M. H., (Adenine Press), Vol. 3, pp. 55–72.
17. Dickerson, R. E., Goodsell, D. S., Kopka, M. L. & Pjura, P. E. (1987) *J. Biomol. Struct. Dynam.* **5**, 557–579.
18. Calladine, C. R. (1982) *J. Mol. Biol.* **161**, 343–352.
19. Ulyanov, N. B. & Zhurkin, V. B. (1984) *J. Biomol. Struct. Dynam.* **2**, 361–385.
20. Srinivasan, A. R., Torres, R., Clark, W. & Olson, W. K. (1987) *J. Biomol. Struct. Dynam.* **5**, 459–496.
21. Mazur, J., Sarai, A. & Jernigan, R. L. (1989) *Biopolymers* **28**, 1223–1233.
22. Sponer, J. & Kypr, J. (1990) *J. Biomol. Struct. Dynam.* **7**, 1211–1220.
23. Lavery, R. & Sklenar, H. (1988) *J. Biomol. Struct. Dynam.* **5**, 655–667.
24. Cambridge Convention for Definition and Nomenclature of Nucleic Acid Parameters, Dickerson, R. E. *et al.* (1989) *J. Mol. Biol.* **205**, 787–789, *EMBO J.* **8**, 1–4, *J. Biomol. Struct. Dynam.* **6**, 627–634.
25. Bhattacharyya, D. & Bansal, M. (1989) *J. Biomol. Struct. Dynam.* **6**, 635–653.
26. Bhattacharyya, D. & Bansal, M. (1991) *J. Biomol. Struct. Dynam.* **8**, 539–571.
27. Arnott, S., Dover, S. D. & Wonacott, A. J. (1969) *Acta Crystallogr.* **B 25**, 2192–2206.
28. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, (Springer-Verlag, New York).
29. Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986) *J. Comp. Chem.* **7**, 230–252.
30. AMBER Ver. 3.0 (1987) Molecular Mechanics Program from University of California, San Francisco.
31. Arnott, S. & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun.* **47**, 1504–1510.
32. Chandrasekaran, R., Wang, M., He, R.-G., Puigjaner, L. C., Byler, M. A. Millane, R. P. & Arnott, S. (1989) *J. Biomol. Struct. Dynam.* **6**, 1189–1202.
33. Arnott, S. & Chandrasekaran, R. (1982) Personal Communication.
34. Bansal, M. & Gupta, G. (1981) *Int. J. Quantum. Chem.* **20**, 407–417.
35. Drew, H. R., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson R. E. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2179–2183.
36. Coll, M., Frederick, C. A., Wang, A. H.-J. & Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8385–8389.
37. Coll, M., Aymami, J., van der Marel, G. A., van Boom, J. H., Rich, A. & Wang, A. H.-J. (1989) *Biochemistry* **28**, 310–320.
38. Yoon, C., Prive, G. G., Goodsell, D. S. & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6332–6336.
39. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Mayer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542.

40. Wang, A. H.-J., Fujii, S., van Boom, H. J., van der Marel, G. A., van Boeckel, S. A. A. & Rich, A. (1982) *Nature (London)* **299**, 601–604.
41. Shakked, Z., Rabinovich, D., Cruse, W. B. T., Egert, E., Kennard, O., Sala, G., Salisbury, S. A. & Viswamitra, M. A. (1981) *Proc. R. Soc. London, Ser. B*, **213**, 479–487.
42. Mohanty, D. & Bansal, M. (1991) *J. Biomol. Struct. Dynam.* **9** (in press).
43. Bansal, M. & Pattabiraman, N. (1989) *Biopolymers* **28**, 531–548.
44. Kitchin, P. A., Klein, V. A., Ryan, K. A., Gann, K. L., Rauch, C. A., Kang, D. S., Wells, R. D. & Englund, P. T. (1986) *J. Biol. Chem.* **261**, 11302–11309.
45. Diekmann, S. & Wang, J. C. (1985) *J. Mol. Biol.* **186**, 1–11.
46. Zahn, K. & Blattner, F. R. (1987) *Science* **236**, 416–422.
47. Ulanovsky, L. & Trifonov, E. N. (1987) *Nature (London)* **326**, 720–722.

Molecular Recognition in Peptides and Proteins

Andrew J. Doig and Dudley H. Williams

Cambridge Centre for Molecular Recognition, University Chemical Laboratories,
Lensfield Road, Cambridge CB2 1EW, England

Introduction

Molecular recognition is found in all areas of biochemistry. It occurs when two molecules bind to one another via non-covalent interactions. The chances of any two randomly chosen molecules forming a bound complex are extremely small, particularly in water where nearly all biological processes take place. If binding is to occur, the surfaces of the two molecules (A and B) must be a good match for each other. This means, for example, that any positively charged group on molecule A must be brought close to a negatively charged group on molecule B, non-polar regions on A must be adjacent to non-polar regions on B and most importantly, the two surfaces must fit snugly so that there are no holes or groups too bulky to fit. In other words, A and B “recognise” each other, like two jigsaw puzzle pieces. This requirement means that molecular recognition is highly specific. It is the basis for the precise control of biochemical reactions, leading, for example, to tight regulation of the concentration of cellular metabolites and accuracy of DNA, RNA and protein synthesis.

Molecular recognition is found everywhere in biochemistry, such as in enzymes which catalyse reactions by binding substrate molecules into their active sites. It is essential that binding is specific, so that the correct molecule undergoes the reaction. The enzyme should also bind strongly to the transition state of the reaction it is catalysing, to increase the reaction rate. An excellent example of molecular recognition is allosterism by which the activity of enzymes is commonly regulated. In an allosteric interaction, regulatory molecules bind to a second site on the enzyme, inducing a conformational change which alters the enzyme's activity. For instance, phosphofructokinase catalyses the transformation of fructose 6-phosphate into fructose 1,6-diphosphate. This is metabolized further, via the biochemical pathway known as glycolysis, to produce a large amount of energy for the cell. The activity of phosphofructokinase is inhibited by high levels of phosphoenolpyruvate, an end product of glycolysis, whose presence shows that there is no need to send more fructose 6-phosphate into the glycolytic pathway. By contrast, high levels of ADP indicate that the cell is short of energy; ADP binds to phosphofructokinase increasing its activity.

Hormones are small molecules which exert large effects, generally by binding to enzymes, switching them on like an allosteric activator and initiating a “cascade” of further reactions. For instance, the hormone insulin is common to species from bacteria to man, and signals the “fed” state. In humans, a measure of the fed state is a high blood glucose level. Insulin promotes the entry of glucose into muscle and fat cells, decreasing the blood glucose level. In contrast, the hormone glucagon signals that glucose is scarce. Both of these hormones exert their effect primarily by binding

to receptors in the liver, leading to a host of further reactions which affect the blood glucose level.

Organisms which lack an immune system (such as plants and bacteria), produce a vast number of natural products, presumably to help them compete with rival species. These secondary metabolites (secondary as they do not lie on main biochemical pathways) must operate by recognising and binding to receptors in the target species (1). One famous secondary metabolite is penicillin (figure 1) which acts by deactivating an enzyme responsible for one of the steps in the synthesis of bacterial cell walls. This enzyme catalyses a reaction between a D-Ala-D-Ala-CO₂⁻ group in the nascent cell wall and a serine residue in the active site of the enzyme, forming a D-Ala-CO-Enzyme intermediate. We believe that penicillin has evolved to bind to this active site and react irreversibly with the serine, rendering the enzyme inactive. However, it must be said that only a minority of secondary metabolites have a known function and of these, very few have a known receptor site.

Despite the diversity of locations in which molecular recognition operates, some common features can be discerned. If we are to understand any area of molecular recognition, it is that exhibiting these common features that is most important: First, the majority of biochemical reactions occur in aqueous solution. The major exceptions to these take place within the liquid crystals of membranes, which are relatively poorly understood. Water is an exceptional solvent in many ways. It is highly polar and bonds to solute molecules much more strongly than a non-polar solvent such as chloroform. This means that the binding of two polar molecules to each other is often much stronger in chloroform than in water, due to the strong solute-water bonds lost on binding. Second, a high proportion of biological molecules are made up of only five types of atoms, namely hydrogen, carbon, oxygen, nitrogen and sulphur. Thus in simplifying our study of molecular recognition we can discount 95% of the periodic table, including all the metals. The particular classes of molecules we have chosen to study are peptides and proteins in water. Peptides are indeed made up of only the five elements listed above but despite this, different amino acids show a wide range of properties. Amino acid side-chains contain polar and non-polar groups, charged and uncharged groups, and aromatic rings and have a range of sizes and flexibility. We shall choose to study two topics in particular: peptide-peptide binding (specifically for the vancomycin class of antibiotics) and protein folding. Both these systems have a wide range of thermodynamic and structural information avail-

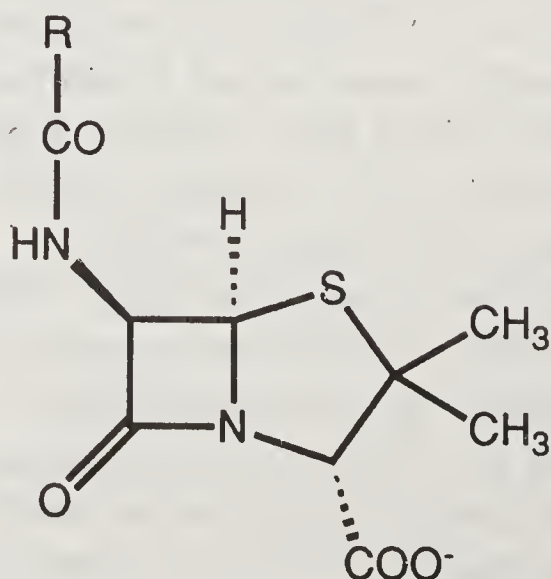


Figure 1. Penicillin.

able. Our model systems for peptide–peptide interactions are the antibiotics vancomycin and ristocetin, both of which bind peptide sequences terminating in D-Ala–D-Ala. By varying the substrate bound by the antibiotic we can determine intrinsic binding energies for different functional group–functional group interactions. For instance, if we compare the binding energies of D-Ala–D-Ala with D-Ala–D-Gly, we can find the contribution of the methyl group in the C-terminal alanine to binding. The effect of each functional group depends on the environment in which it is placed in the antibiotic. We hope that the results will be generally applicable to all peptide–peptide interactions (including proteins). Ultimately, we aim to predict binding constants which would be of obvious benefit in the rational design of drugs. Several hundred protein structures have been determined, principally by X-ray crystallography, though nuclear magnetic resonance (NMR) looks set to play a major role in determining protein structure in the future. We have utilised known protein structures, together with thermodynamic data, to investigate the properties of proteins, including their surface areas and shapes when they unfold thermally and the effects of disulphide bridges on the thermodynamics of protein folding.

Molecular Recognition in the Vancomycin Group of Antibiotics

The Structure Elucidation of Vancomycin and Ristocetin

For a detailed analysis of molecular recognition it is necessary to know the structures of both substrate and receptor. In this section we describe briefly the structure elucidation of the vancomycin group. Knowledge of this receptor structure leads to an understanding of the molecular basis of action. Vancomycin is a glycopeptide antibiotic obtained from *Streptomyces orientalis* and it is bactericidal against Gram-positive bacteria. It was first isolated in 1956 (2), and now enjoys widespread clinical use, especially in the treatment of severe staphylococcal infections (e.g. endocarditis and wound septicaemia). A recent development has been its use orally in the treatment of pseudomonas colitis, a severe and potentially lethal disease associated with antibiotic treatment after major gastrointestinal surgery (3). The structure elucidation of vancomycin began in the 1960s. At that time suitable methods for this task were not available and workers had to be content with the identification of chemical degradation products. These included N-methyllucine, aspartic acid, glucose and chlorophenols (4, 5), each of which was subsequently found as a part of the antibiotic. The advent of high field NMR in the 1970s meant it was possible to fully assign the ^1H NMR spectrum of vancomycin (6), in terms of all the structural units which had been isolated via chemical degradation (7–10). These units could be assembled into a partial structure by utilising the nuclear Overhauser effect (nOe), which gives information on the spatial proximity of protons. The next important step for the structure elucidation of the vancomycin group was the determination of the X-ray structure (11) of a degradation product of vancomycin called CDP-I (4). The structure of CDP-I could be combined with the known degradation products and their sites, to give the vancomycin structure. However, further nOe studies showed that this structure was in error, as one aromatic ring had rotated by 180° in the conversion from vancomycin to CDP-I (12). The correct structure of vancomycin is shown in figure 2. Following this work, the structures of a number of other antibiotics

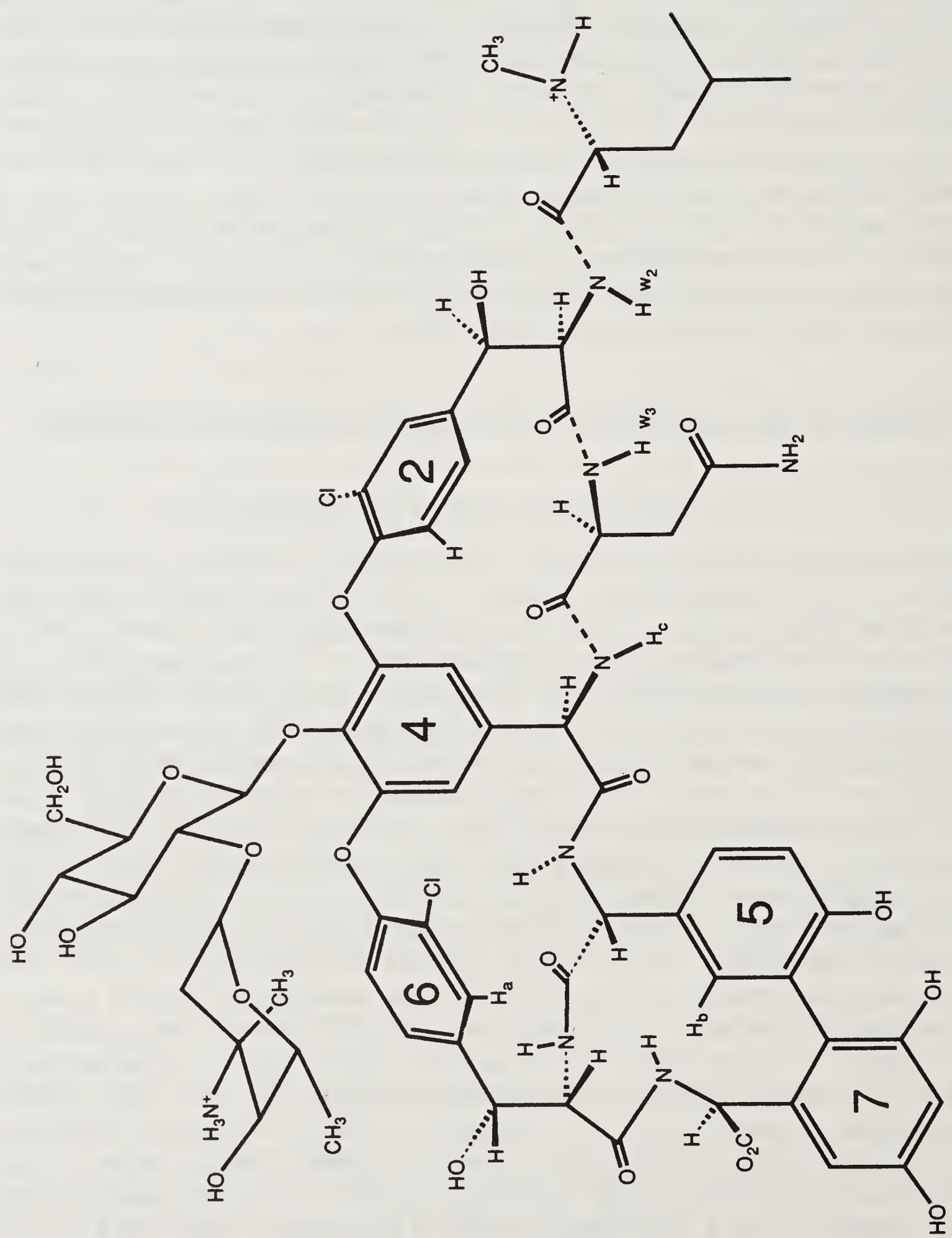


Figure 2. Vancomycin.

in the group have been determined. These include ristocetin (13–15), avoparcin (16,17) antibiotic A35512 (18–20) and actaplanin (A-4696; 21, 22). All are linear heptapeptides with unusual amino acids linked by phenolic oxidative couplings. The structure of ristocetin A is given in figure 3. This is dissimilar to vancomycin in that all the amino acid side chains are cross-linked. All the compounds share identical oxygen substitution patterns on rings 4, 5, 6 and 7. The stereochemistry at the asymmetric centres associated with the seven amino acids is conserved and from the N-terminus is R, R, S, R, R, S, S.

Biosynthetic studies have established that both the β -hydroxychlorotyrosine units of vancomycin and both the β -hydroxytyrosine units of ristocetin can be derived from tyrosine (23, 24). It was also established that the *p*-hydroxyphenylglycine units (numbering amino acid residues 1–7 from the N- to the C-terminus) at positions 4 and 5 in vancomycin and at positions 1, 4 and 5 in ristocetin A can be derived from tyrosine. The *m*-dihydroxylated phenylglycine units (residue 7 in vancomycin and residues 3 and 7 in ristocetin) are derived from acetate. The aromatic methyl group of residue 3 in ristocetin A is presumably derived from *S*-adenosylmethionine, but this has not been established experimentally. The polycyclic structures are formally completed from a heptapeptide by three phenol oxidative couplings in vancomycin

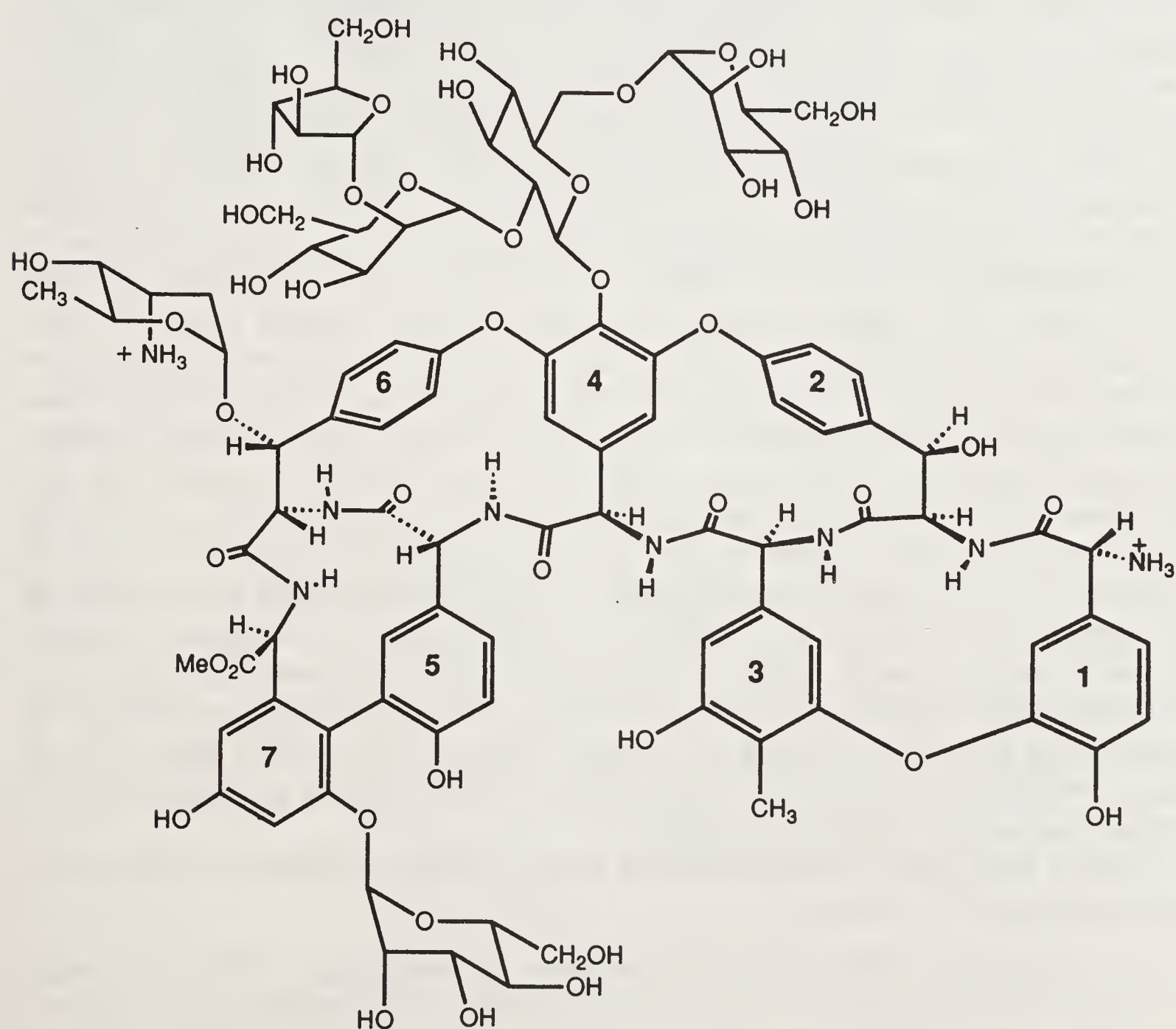


Figure 3. Ristocetin.

and four in the case of ristocetin A, followed by the addition of the sugars. The order of these steps has not been established.

The Mode of Action of the Vancomycin Group

Like penicillin, the vancomycin group of antibiotics inhibit bacterial cell wall biosynthesis. Vancomycin acts by binding to cell wall precursors ending in the peptide D-Ala-D-Ala (25). The molecular basis of action was probed by studying proton NMR spectra of the cell wall analogue N-Ac-D-Ala-D-Ala, the isolated antibiotic and the N-Ac-D-Ala-D-Ala/vancomycin complex. The first study (26) showed that the C-terminal alanine methyl resonance of N-Ac-D-Ala-D-Ala was shifted to high field by ~ 0.9 ppm upon complexation. It was concluded that this methyl lay over the face of a benzene ring of the antibiotic in the complex. Later ^1H NMR work also reported structural details of a 1:1 mixture of vancomycin and N-Ac-D-Ala-D-Ala (6). The complex was in fast exchange with the free components at room temperature in d_6 -DMSO, and it was possible to follow the averaged shifts of a number of antibiotic protons on complex formation. In particular, protons H_b and H_c moved downfield and H_a moved upfield (figure 2). It was therefore concluded that all these protons lay close to the bound dipeptide in the complex and that H_c is hydrogen-bonded to a dipeptide carbonyl group. This carbonyl is likely to be that of the carboxy terminus of the peptide, since when this interacts with NH_c , the methyl group of the C-terminal D-alanine can be over benzene ring 4, thus giving rise to the observed upfield shift (26). Before any further information on the mode of action of vancomycin was obtained, a more complete binding site for ristocetin was deduced (27). This was possible because the ristocetin/N-Ac-D-Ala-D-Ala complex is in slow exchange with the free species, on the ^1H NMR timescale. This meant that the chemical shifts of many protons (including all the NH protons of the secondary amides in both components) could be observed in both the free and bound states. All those NH protons which underwent large downfield shifts on complex formation were involved in hydrogen-bond formation. Second, protons which were hidden from solvent in the complex showed a low temperature dependence of chemical shift. Those which remained solvent-exposed showed a large movement of chemical shift to higher field with increasing temperature. Analysis of these results indicated that the C-terminal alanine amide proton is hidden from solvent after N-Ac-D-Ala-D-Ala binds, but that of the N-terminal alanine is still exposed. The six amide NH's of ristocetin A were divided into two groups: two NH protons which were located on the convex "back" face of the antibiotic remained exposed to solvent on complex formation and four amide NH's, located within a cleft in the molecule, took part in hydrogen-bond formation on complex formation. These experiments established the geometry of the antibiotic/dipeptide complex and gave, with some help from the earlier work on the vancomycin mode of action, a model for the molecular basis for complex formation in ristocetin (figure 4).

Figure 5 shows the hydrogen-bonding in the ristocetin-dipeptide complex. The main features of the complex are:

- (i) A carboxylate binding pocket in ristocetin A, consisting of NH's 2, 3 and 4 (marked w_2 , w_3 and w_4 in figure 4), which are in a cleft flanked by hydrophobic walls constituted from the aromatic rings of residues 1 and 2; the carboxyl group

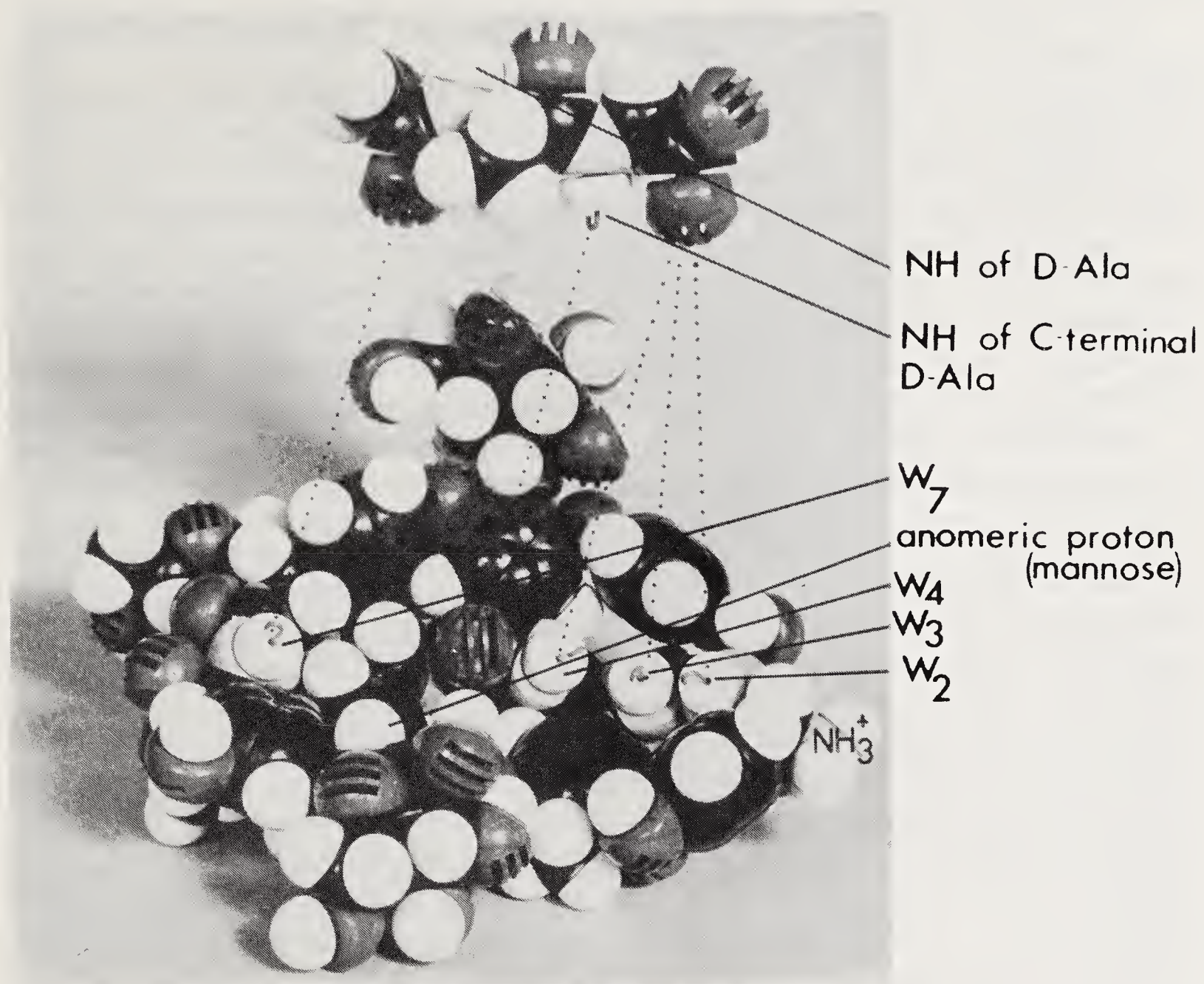


Figure 4. Interactions formed in the ristocetin/dipeptide complex.

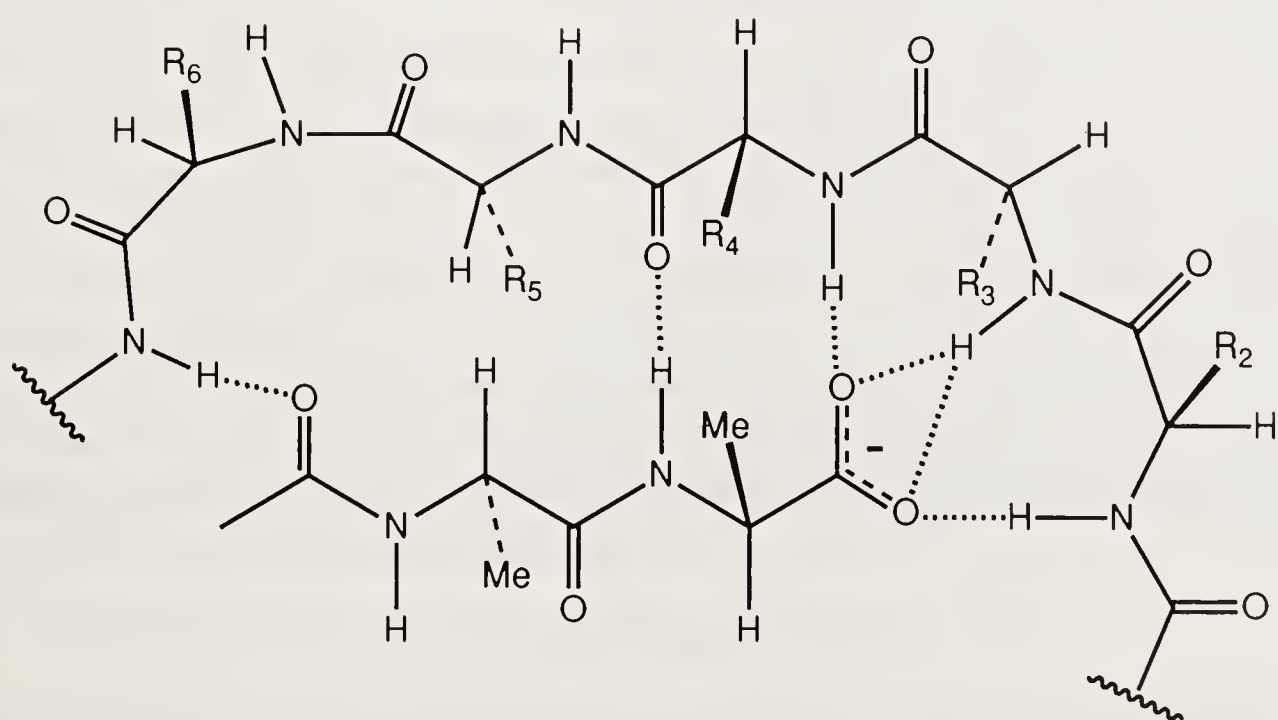


Figure 5. Hydrogen-bonding in the ristocetin/dipeptide complex.

- of N-Ac-D-Ala-D-Ala binds into this pocket by hydrogen-bond formation between the NH groups 2, 3 and 4.
- (ii) Binding of the carboxylate anion is slightly strengthened by a favourable Coulombic attraction between the CO_2^- of the cell wall analogue and the NH_3^+ at the N-terminus of the antibiotic.
 - (iii) The NH of the C-terminal D-alanine of the dipeptide forms a hydrogen-bond to the carbonyl group of residue 4 of the antibiotic.
 - (iv) NH-7 of the antibiotic (marked w_7 in figure 4) forms a hydrogen-bond to the carbonyl oxygen of the N-acetyl group of the dipeptide.
 - (v) When the above five hydrogen-bonds are made, the methyl group of the C-terminal alanine forms a hydrophobic bond with the face of the aromatic ring of residue 4 and the methyl group of the other alanine forms a hydrophobic bond with the benzene rings of residues 5 and 7.

All these five features contribute towards binding. Our present aim is to quantitatively understand the forces involved, via the determination of intrinsic binding free energies, enthalpies and entropies. The ristocetin system is ideal for this purpose because the antibiotic is relatively rigid, due to extensive cross-linking of side chains. This means that there is little complication caused by conformational equilibria in unbound antibiotic. Second, the structure of the complex and free molecules is known in great detail (28–30) mainly through the use of the nOe. Thirdly, ristocetin binds very simple molecules. This means that structural analogues of the cell wall peptide are readily bought or synthesized and a large amount of data can be collected quickly. This contrasts with the large amount of time and effort required to produce mutant proteins. Also, the smaller the molecule, the better are our chances of determining its unbound conformation. Before discussing what we have learnt on peptide–peptide binding in the ristocetin system, we will first examine the forces involved in molecular recognition.

Thermodynamics of Peptide–Peptide Binding

Translational and Rotational Entropies

The general reaction $A + B \rightarrow AB$, where AB is the bound complex does not occur unless there are many favourable forces driving the binding. Why is this? The major problem is that both A and B have large translational (S_{trans}) and rotational (S_{rot}) entropies in solution. These translational and rotational entropies have little dependence on mass (or moment of inertia), so that the complex AB has translational and rotational entropy not very different to those of the larger of A and B. This means that a large amount of translational and rotational entropy is lost in bimolecular association. To overcome this loss of translational and rotational entropy, there need to be many favourable interactions within the complex (such as hydrogen-bonds, salt-bridges, hydrophobic interactions etc.) In the gas phase S_{trans} is given for a perfect gas by the Sackur–Tetrode equation:

$$S_{\text{trans}} = 5R/2 + (5R/2)\ln T - R \ln P + R \ln [(2\pi m/h^2)^{3/2} k^{5/2}], \text{ JK}^{-1} \text{ mol}^{-1}. \quad (1)$$

S_{trans} depends on the $\ln(m)$, which in practice means that there is little variation of translational entropy with mass. For example, in the gas phase $S_{\text{trans}} = 204 \text{ JK}^{-1} \text{ mol}^{-1}$

for ristocetin A, $180 \text{ JK}^{-1} \text{ mol}^{-1}$ for N-Ac-D-Ala-D-Ala and $205 \text{ JK}^{-1} \text{ mol}^{-1}$ for their bound complex. S_{rot} for a gas, is given by equation (2):

$$S_{\text{rot}} = R + R \ln [\pi^{7/2} (8kT/h^2)^{3/2} (I_A I_B I_C)^{1/2}], \text{ JK}^{-1} \text{ mol}^{-1}. \quad (2)$$

The rotational entropy thus depends on the $\ln (I_A I_B I_C)$ (where I_A, I_B and I_C are the moments of inertia about three perpendicular axes). As with translation, this means that there is relatively little increase of rotational entropy as the size of the molecule increases. For ristocetin A in the gas phase, $S_{\text{rot}} = 188 \text{ JK}^{-1} \text{ mol}^{-1}$, for N-Ac-D-Ala-D-Ala $S_{\text{rot}} = 138 \text{ JK}^{-1} \text{ mol}^{-1}$ and for their complex $S_{\text{rot}} = 191 \text{ JK}^{-1} \text{ mol}^{-1}$. Analysing all these entropies we see that the bimolecular association must overcome unfavourable translational and rotational entropies of $314 \text{ JK}^{-1} \text{ mol}^{-1}$. ($\Delta S = S_{\text{trans}}$ (ristocetin) + S_{trans} (dipeptide) + S_{rot} (ristocetin) + S_{rot} (dipeptide) - S_{trans} (complex) - S_{rot} (complex) = $314 \text{ JK}^{-1} \text{ mol}^{-1}$.) There is a small enthalpy contribution of 10 kJmol^{-1} which altogether gives an unfavourable free energy change for binding of 84 kJmol^{-1} (obtained using $\Delta G = \Delta H - T\Delta S$). The above equations are reasonably accurate for molecules in the gas phase. However, biochemical molecular recognition occurs in an aqueous environment. Since we know the gas phase entropies, we will start from there and consider how they change when the molecules in question are transferred to solution in water. We can break this down into two stages: first, transfer from gas to pure liquid and second, dilution of the liquid with water. Entropies of vaporisation (transfer from liquid to gas phase) can be predicted for non-polar molecules using the Hildebrand rule (31). This gives the entropy of vaporisation as a function of the boiling point of the molecule. However, this is unsuitable for our peptides because we cannot know their boiling points. Even the simplest peptides, such as alanine and glycine, decompose on melting. They are also polar so the Hildebrand rule will be inaccurate. Entropies of vaporisation are commonly determined by measuring the vapour pressure above the liquid, and applying the Clausius–Clapeyron equation (32). This approach will not work in our system as we cannot obtain a liquid peptide. Therefore, we can only guess the entropy of vaporisation. For the dilution step we can use the term in the Sackur–Tetrode equation (1), which depends on pressure. If we consider the peptide only, we can regard the dilution as a drop in pressure, since each peptide molecule has a greater volume which it can occupy. Finally, water will bind to the peptide. Bound waters will increase the effective mass and moment of inertia of the peptide, thus increasing S_{trans} and S_{rot} (equations 1 and 2). It has been pointed out that, in some cases, translational and rotational entropies are similar for liquids and gases, e.g. cyclopentadiene and its dimerisation product (33). On the other hand, the experimental difference in entropies of numerous simple gases and liquids gives a value ($S_{\text{gas}} - S_{\text{liquid}}$) of $\approx 80 \text{ JK}^{-1} \text{ mol}^{-1}$ (Trouton's rule). It is impossible at present to calculate translational and rotational entropies of peptides in aqueous solution. We therefore aim to measure them. Consider binding a large peptide A to ristocetin and measuring its free energy of binding. If we can divide A into two pieces B and C, both of which also bind to ristocetin, the sum of the binding energies of B and C differ from the binding energy of A by the difference in the translational and rotational energies.

Complementarity

As mentioned in the introduction, molecular recognition depends on a complementarity of shape between the interacting molecules. A receptor and substrate

Table 1. The effect on disrupting complementarity on binding energy (taken from Fersht 1985).

Binding cavity		Unfavourable binding Energy (kJmol ⁻¹)
Constructed for	Occupied by	
-CH ₃	-OH	14
-H	-CH ₃	> 32
-H	-OH	15
-OH	-H	29
-NH ₃ ⁺	-H	18

can be regarded, for a first approximation, as a lock and key, which fit perfectly leaving no gaps between their packing surfaces. All hydrogen-bonds and salt linkages are paired. The effect of disrupting complementarity within proteins has been investigated by Fersht (34–36) in the enzyme aminoacyl-tRNA synthetase (table 1). These large changes in binding energy show the magnitudes of the forces that give rise to specificity in binding. The effects of deleting methylene (CH₂) groups from the hydrophobic core of proteins have also been measured (37) and analysed (38). The essential requirement of complementarity means that the design of two molecules which will bind to one another is very difficult. However, natural products have evolved so that, where so far examined, the complementarity between receptor and substrate is good. This problem is therefore not present when studying natural receptors and natural substrates.

Non-Covalent Bonds

All bonds are electrostatic in origin – that is they depend on attraction between opposite charges. It is useful to divide non-covalent bonds into several categories. In this section we will consider electrostatic interactions between charged ions or dipoles (molecules with unevenly distributed charge), hydrogen-bonds and van der Waals interactions.

The most familiar physical interaction is the electrostatic force that occurs between any two charged particles (such as a salt-bridge -NH₃⁺ -O₂C-). The magnitude of this interaction in solution is given to a first approximation by Coulomb's law (equation 3):

$$\Delta E = Z_A Z_B \epsilon^2 / D r_{AB}, \quad (3)$$

where Z_A and Z_B are the charges on the two particles, ϵ is one unit of electronic charge, D is the dielectric constant for the solvent and r_{AB} is the distance between the two particles. In a vacuum $D = 1$, in non-polar solvents $D = 2$ –10 and in water $D = 80$. This means that electrostatic bonds in water are much weaker than those in a vacuum. Note that the interaction energy falls off with $1/D r_{AB}$; this means that electrostatic forces can be relatively long range. Electrostatic bonds involving dipoles are shorter range: between orientated dipoles the energy falls off with $1/D r_{AB}^3$, between an ion and an ion-induced dipole the energy falls off with $1/D^2 r_{AB}^4$ and between a permanent dipole and a dipole-induced dipole the energy falls off with $1/D^2 r_{AB}^6$. These latter forces are significant when the interacting atoms are very close; this

is perhaps why the insertion of a group which is too large into a flexible binding site is so unfavourable – atoms can no longer maintain close contacts and these bonds effectively are lost or reduced in strength. The dielectric constant within a binding site may be less than that in water, to which the substrate was previously exposed; this will strengthen the electrostatic bonds formed in a non-polar environment. For example, a salt-bridge in the middle of the hydrophobic core of a protein will be stronger than a salt-bridge between identical functional groups in water.

A special case of an electrostatic interaction is the hydrogen-bond. This consists of two electronegative atoms bound to the same hydrogen. One particularly good hydrogen-bond donor is the NH of a polypeptide backbone; one excellent acceptor is the C=O group in the backbone. Hydrogen-bonds between these two donors and acceptors are the basis of the two major units of protein secondary structure, namely the α -helix and the β -sheet. Hydrogen-bonds are strongest when the participating atoms lie in a straight line; the exact relationship of bond strength to bond angle is not known (39).

The weakest interaction between atoms is the van der Waals or London dispersion force. All atoms are subject to these interactions, indeed it is the only interaction between noble gases. The weakness of the force leads to the very low melting point of 4 K for helium. Van der Waals forces arise from an unequal instantaneous distribution of electrons about the nucleus. These transient dipoles induce dipoles in neighbouring atoms producing a weak bond. Van der Waals forces follow an inverse sixth power law. At close distances the atoms repel, approximately following an inverse twelfth power law. This corresponds to the Lennard–Jones potential energy function for van der Waals forces (equation 4):

$$\Delta E = -A/r^6 + B/r^{12}, \quad (4)$$

where A is a constant, B is a constant which depends on the polarisability of the atoms in question and r is the distance between the interacting atoms. This equation has a minimum energy distance, known as the van der Waals contact radius. Though van der Waals energies are low, they are additive and can sum to significant values. In a crystalline hydrocarbon, each methylene group has 8.4 kJmol^{-1} of van der Waals energy and each methine group 6.7 kJmol^{-1} (36).

The Hydrophobic Effect

The hydrophobic effect can be summarised as “oil and water do not mix”. In other words non-polar molecules are observed to interact with one another rather than with water. The hydrophobic effect can be measured, as a free energy, using solvent transfer experiments, where the non-polar molecule is partitioned between water and an organic solvent (41–45). The transfer of a non-polar molecule from organic solvent to water can be used as a model for non-polar groups in a protein upon protein unfolding (46). These solvent transfer experiments, where hydrocarbon is transferred from liquid hydrocarbon to water, reveal the following properties of the hydrophobic effect:

- (i) The free energy for the transfer is large and positive at all temperatures, meaning that the transfer is always highly unfavourable.
- (ii) The change in heat capacity (ΔC_p) is large and positive at all temperatures which leads to a large change in enthalpy and entropy with temperature. In protein

- unfolding, the hydrophobic effect is believed to be the sole contributor to the change in heat capacity (see refs. 47–50 for discussion).
- (iii) At room temperature the enthalpy change is near to zero and the entropy change is negative. This means that at room temperature the hydrophobic effect is entropy driven.
 - (iv) At high temperature (110°C), the entropy change is approximately zero and the enthalpy change is positive. In other words, as the temperature increases from room temperature, the hydrophobic effect remains large but becomes enthalpy, rather than entropy, driven.
 - (v) The free energy change for the transfer is directly proportional to the water accessible surface area of the non-polar group (A_{np}) (51–53). This surface is obtained by rolling a sphere, the size of a water molecule (radius 1.4 Å), over the hydrocarbon. This is a reasonable model because the water accessible surface area reflects the number of water molecules which are able to solvate the non-polar group.
 - (vi) The free energy change for the transfer is directly proportional to the change in heat capacity.
 - (vii) The change in water accessible non-polar surface area (ΔA_{np}) is directly proportional to the change in heat capacity (47, 54–57). The size of the hydrophobic effect can be calculated if we know the structures of the free species and bound complex, by measuring the water accessible non-polar surface area.

Internal Rotations

Internal rotations are rotations about single bonds within a molecule, rather than the tumbling of the molecule as a whole in space (discussed above). Their theoretical contribution to enthalpy and entropy has been determined by Pitzer (58). In order to calculate the effect of an internal rotor, we need to know the magnitude of the potential barrier for the rotation and the moments of inertia of the two groups on either end of the bond about which the rotation is taking place. For simple molecules which are not interacting with solvent, such as ethane in the gas phase, this is straightforward. For these small molecules the entropy of the internal rotor is 13 to 35 JK⁻¹mol⁻¹, its enthalpy is 1.3 to 2.5 kJmol⁻¹, giving a free energy of –2.6 to –7.9 kJmol⁻¹ at room temperature. Note that the entropic contribution to the free energy is much greater than the enthalpic contribution. This is the free energy of the rotation; when the substrate binds, and the rotor is “frozen out”, the rotor is replaced by a torsional vibration. This will have a free energy of $\cong 2$ kJmol⁻¹, assuming a torsional vibration frequency of $\cong 500$ cm⁻¹, which leads to a change in free energy when a rotor is frozen out of 0 to 6 kJmol⁻¹. The rotors in Ac-D-Ala-D-Ala will each contribute about 6 kJmol⁻¹ in free energy opposing binding to the antibiotic. Water molecules bound to polar peptides again present a problem. These will increase the moments of inertia and increase the free energy of each rotor. The rate of rotation of species in polar solvents decreases with increasing charge, because of greater bonding to solvent molecules (29). Further problems are encountered as the size of the molecule increases. First, the molecule will exist in a number of conformations, making the calculation of the moments of inertia difficult. Second, internal rotations will become cooperative, that is, several bonds will rotate simultaneously, avoiding large motions of the ends of the chain. This will decrease the moments of inertia about the rotating bond. Since calculation of the effects of internal rotations is too

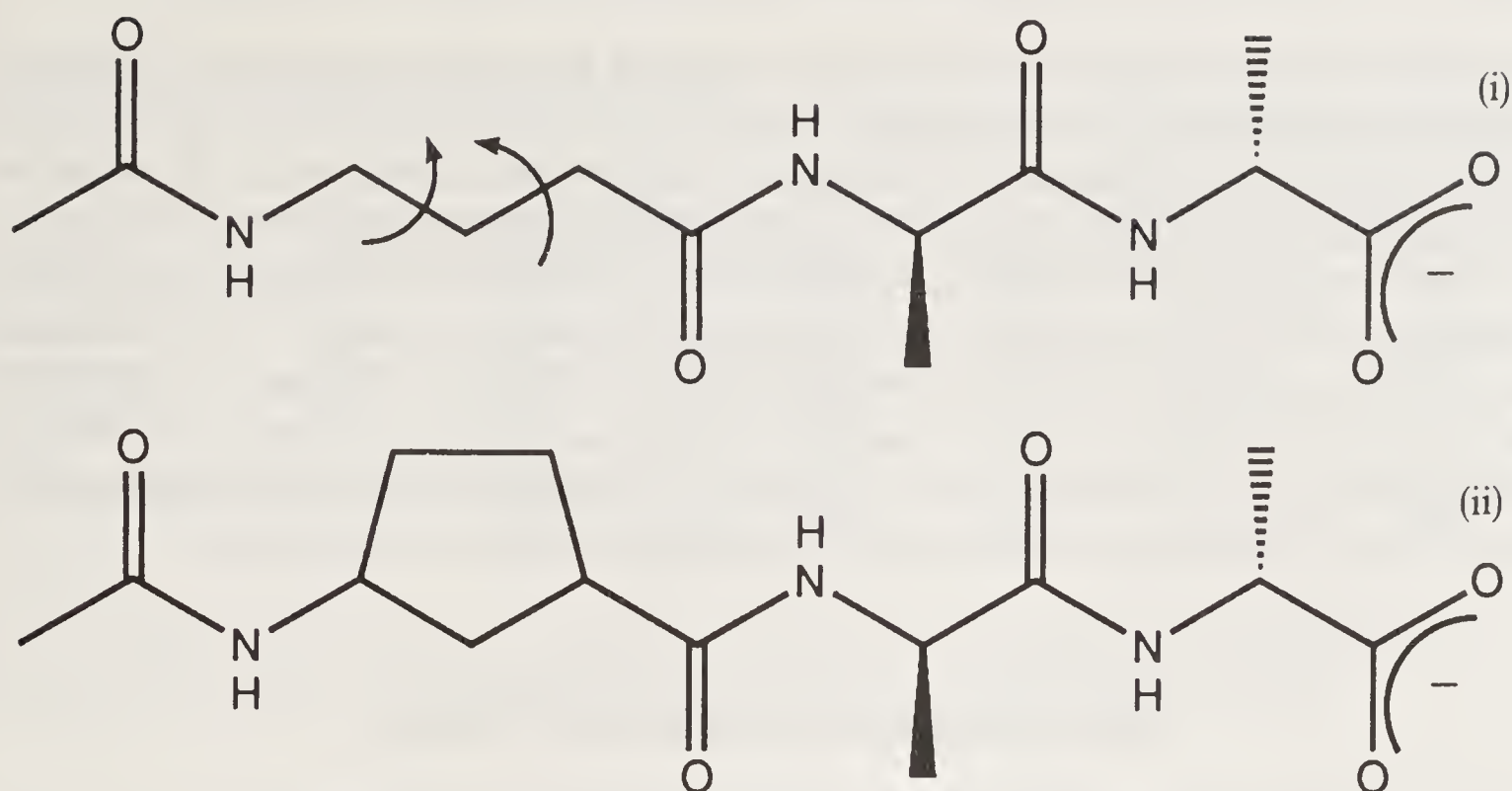


Figure 6. Peptides which differ by two internal rotations.

difficult, we aim to measure them. Figure 6 shows two peptides, both of which are expected to bind tightly to ristocetin. They differ in that peptide 6(ii) has an extra two methylene groups which form a five-membered ring. The presence of this ring means that rotation about the two bonds marked in peptide 6(i) is impossible. The additional methylenes form no contacts with the ristocetin, so they remain solvated by water when free and when bound. Peptide 6(ii) is expected to bind to ristocetin more strongly than peptide 6(i) since an additional two internal rotors must be “frozen out” in binding peptide 6(i). The difference in binding energies is equivalent to the contribution of the two internal rotors, marked on peptide 6(i), to the free energy of unbound 6(i).

Thermodynamics of Peptide–Peptide Binding in Ristocetin and Vancomycin

The ristocetin system is ideal for the investigation of the thermodynamics of molecular recognition, in particular for binding between peptides. In analysing this system it is essential to consider the solvent water. When peptide and antibiotic bind many water molecules are released. These were bound to the groups involved in hydrogen-bonding in the complex or those solvating non-polar groups buried in the complex. Ristocetin is a chiral molecule and therefore shows stereoselectivity. For example, the binding energy of $\text{C}_6\text{H}_5\text{CO-D-Ala-D-Ala}$ is $15\text{--}21\text{ kJmol}^{-1}$ more than that of $\text{C}_6\text{H}_5\text{CO-D-Ala-L-Ala}$ (59). Ristocetin is therefore able to discriminate between a D and an L alanine by a factor of 500–4000. Intrinsic binding constants for amide–amide hydrogen-bonds are available from a comparison of the calorimetric data of Rodriguez-Tebar *et al.* (60) on the binding of ristocetin and vancomycin. Binding of the two peptide substrates N-Ac-D-Ala and N-Ac-Gly-D-Ala differs only by the formation of an extra hydrogen-bond in the binding of the dipeptide. The difference in binding enthalpy of the two peptides is $-3.4 \pm 2.8\text{ kJmol}^{-1}$ for ristocetin and $1.4 \pm 2.3\text{ kJmol}^{-1}$ for vancomycin. Both these values are small. We conclude that in an aqueous environment, the hydrogen-bond strength between amide and amide, adjusted for the unfavourable enthalpy change associated with the release of water molecules from the carbonyl and amide units involved, is approximately zero. Calorimetric data (60) show that the entropy change associated with formation of

the same hydrogen-bond, and concurrent freezing out of the two rotors of glycine necessary to make this hydrogen-bond, is (in terms of $-T\Delta S$ at 298 K) $11 \pm 3 \text{ kJmol}^{-1}$ in the case of vancomycin and $8 \pm 4 \text{ kJmol}^{-1}$ in the case of ristocetin. We calculate that freezing out of these two rotors is entropically unfavourable by $\cong 12 \text{ kJmol}^{-1}$ at room temperature (discussed above); therefore the total favourable entropy term associated with the formation of the hydrogen-bond is $16\text{--}24 \text{ kJmol}^{-1}$. We conclude that in these systems (and by extrapolation in isolated β -sheets in general), amide–amide hydrogen-bond formation in aqueous solution is largely an entropy driven process; the favourable entropy change is associated with water release from the CO and NH groups which form the hydrogen-bond in the complex.

Thermodynamics of Protein Folding

The Hydrophobic Effect and the Unfolded State of Proteins

Protein folding is one of the great unsolved problems of contemporary science. Although progress has been made towards solving this problem during the last two decades, it is still not possible in the general case to take a determined primary sequence of amino acids and from this to predict the tertiary structure of the corresponding protein. The challenge to solve this problem is not only intellectually important, but also of great practical relevance in view of the rapidity with which c-DNA sequences may now be obtained, and the corresponding primary amino acid sequences thereby derived.

All the forces and effects discussed above for peptide–peptide binding are relevant to protein folding. The hydrophobic effect is of particular importance (40, 46, 47). All globular proteins are stabilised by burying non-polar side chains (those of valine, leucine, isoleucine, phenylalanine and methionine) within a hydrophobic core and away from water. As discussed previously, we can model the properties of these non-polar side chains in protein unfolding by solvent transfer experiments where hydrocarbons are transferred from liquid hydrocarbon to water. When a protein unfolds there is a large change in heat capacity (ΔC_p) which can be attributed solely to the hydrophobic effect. Since this heat capacity change is directly proportional to the difference in water accessible surface area between the unfolded and folded states (ΔA_{np}), it is possible to calculate this change in surface area. These calculations have been performed for eight proteins for which we have heat capacities and crystal structures (table 2), namely lysozyme, α -chymotrypsin, myoglobin, ribonuclease, parvalbumin, papain, carbonic anhydrase and *staphylococcal* nuclease. The exposed non-polar surface area of the folded protein ($A_{np, \text{folded}}$) is measured from the crystal structure using the MacroModel program (61). The non-polar surface area of the unfolded protein ($A_{np, \text{unfolded}}$) is calculated by adding $A_{np, \text{folded}}$ to ΔA_{np} . The number of residues (n_r) and number of disulphide bonds (n_{ss}) is included, together with the surface area per residue when folded ($A_{np, \text{folded}}/n_r$) and when unfolded ($A_{np, \text{unfolded}}/n_r$). The structure of myoglobin is complicated by the presence of a heme group in the folded protein. It is assumed that the heme fully dissociates when myoglobin unfolds. Table 2 shows that myoglobin has a larger value of $A_{np, \text{unfolded}}$ than the other seven proteins. This is because the other proteins each have a number of disulphide bridges or coordinated metals. These constrain the unfolded proteins, making them much

Table 2. Non-polar surface areas of proteins.

Protein	n_r	n_{ss}	ΔCp	ΔA_{np}	$A_{np, \text{folded}}$	$A_{np, \text{unfolded}}$	$A_{np, \text{folded}}/n_r$	$A_{np, \text{unfolded}}/n_r$
Lysozyme	129	4	6670	4270	2676.5	6950	20.7	53.9
α -Chymotrypsin	245	5	14100	9040	4499.0	13540	18.4	55.3
Myoglobin	153	0	11400	7300	3557.0	10860	23.2	71.0*
Ribonuclease	124	4	5390	3450	3113.1	6560	25.1	52.9
Parvalbumin	108	0	4970	3180	2938.9	6120	27.2	56.7
Papain	212	3	12700	8130	3801.0	11930	17.9	56.3
Carbonic anhydrase	256	0	16200	10400	4892.0	15290	19.1	59.7
<i>Staphylococcal</i> nuclease	141	0	8640	5530	3920.5	9450	27.8	67.0

ΔCp is in $\text{JK}^{-1} \text{mol}^{-1}$. All A_{np} values are in \AA^2 . All A_{np}/n_r values are in $\text{\AA}^2 (\text{mol. res.})^{-1}$.

* Without the heme group. All other myoglobin figures include the heme group.

more compact (decreasing the surface area). Myoglobin has no disulphide bridges, is much more open and flexible in the unfolded state and therefore has a greater proportion of non-polar groups solvated by water. This is the cause of a disproportionately large value of ΔCp .

We can use the value of $A_{np, \text{unfolded}}$ to investigate the geometry of the unfolded protein. In the unfolded state, the protein undoubtedly exists as a vast number of conformers, which are rapidly interconverting (62). Each amino acid has preferred bond angles about the peptide bond within the peptide backbone (63). The backbone conformation of a polypeptide can be described by two dihedral angles per amino acid (ϕ and ψ). In general, there are two pairs of dihedral angles which give lowest energy conformations. One of these conformations gives rise to the α -helix, for several sequential amino acids. The second gives rise to β -sheets; in it two or more β -strands (linear chains of amino acids with β -sheet dihedral angles) are hydrogen-bonded. Each amino acid, whether it is in the folded or unfolded state, is generally found in one of these two conformations. The sequence of amino acids in myoglobin was placed in the anti-parallel β -strand geometry ($\phi = -139^\circ$, $\psi = -135^\circ$) in a polymer with the side chains extended ($\chi_1 = -120^\circ$, $\chi_{n+1} = 180^\circ$). Myoglobin was chosen as it has no disulphides and has the largest value of $A_{np, \text{unfolded}}/n_r$ (table 2), indicating it has unfolded to the largest extent. In this conformation the non-polar surface area ($A_{np, \beta}$) was found to be $11\,500 \text{\AA}^2$ (including the heme group) using MacroModel (61). The β -strand is expected to expose as many side chains to water as possible, since adjacent side chains point in opposite directions and the backbone is in a straight line (except for kinks introduced by prolines). This geometry therefore gives a more physically realistic maximum value of A_{np} than one obtained by summing the surface areas of individual amino acids in a Gly-X-Gly tripeptide (as done by Lee and Richards, ref. 64). Similarly, the non-polar surface area of the sequence of amino acids in myoglobin was measured with the chain in an α -helical conformation ($\phi = -52^\circ$, $\psi = -53^\circ$). $A_{np, \alpha}$ was found to be 8900\AA^2 . This value is smaller than the non-polar surface area for myoglobin in the β -strand since side chains are brought closer together in an α -helix, removing them from contact with water. If we compare the non-polar surface area of unfolded myoglobin derived from experiment ($A_{np, \text{unfolded}} = 10860 \text{\AA}^2$, table 2) we see that the observed surface area is intermediate between $A_{np, \beta}$ and $A_{np, \alpha}$. A surface area in agreement with the observed value could be generated if 75% of the residue in unfolded myoglobin had the dihedral angles of a β -strand and 25%

had the dihedral angles of an α -helix. It should be noted that this interpretation of the results does not imply that numerous adjacent amino acids have β -strand or α -helix geometry, and therefore it is not implied that extended units of secondary structure remain in the unfolded state. Instead, on average, 25% of the residues in myoglobin may have the dihedral angles of an α -helix and 75% have the dihedral angles of a β -sheet. There is no evidence for any residual tertiary structure. Unfolded myoglobin can therefore be well described as a "random coil". The presence of cross-linking disulphide bonds complicates this picture.

Disulphide Bonds in Proteins

Thermodynamic data for protein unfolding are available for 12 proteins (50). These data comprise values of ΔG , ΔH , ΔS and ΔC_p for unfolding at room temperature (298 K). Six of these proteins, namely lysozyme, α -chymotrypsin, papain, ribonuclease, pepsinogen and fragment K4 of plasminogen are monomeric globular proteins with no metals or cofactors in the folded protein. We can use the thermodynamic data for these six proteins to analyse the effect of disulphide bonds on protein stability. If we plot the change in heat capacity on unfolding per residue ($\Delta C_p/n_r$) against the number of disulphides per residue (n_{ss}/n_r) we obtain a straight line (figure 7(i)). By a least squares analysis the data points can be fitted to the equation:

$$\Delta C_p/n_r = 77.3 - 936 n_{ss}/n_r \text{ (JK}^{-1}\text{mol}^{-1}\text{)}. \quad (5)$$

The correlation coefficient (R) is 0.92. Multiplying by n_r gives:

$$\Delta C_p = 77.3 n_r - 936 n_{ss} \text{ (JK}^{-1}\text{mol}^{-1}\text{)}. \quad (6)$$

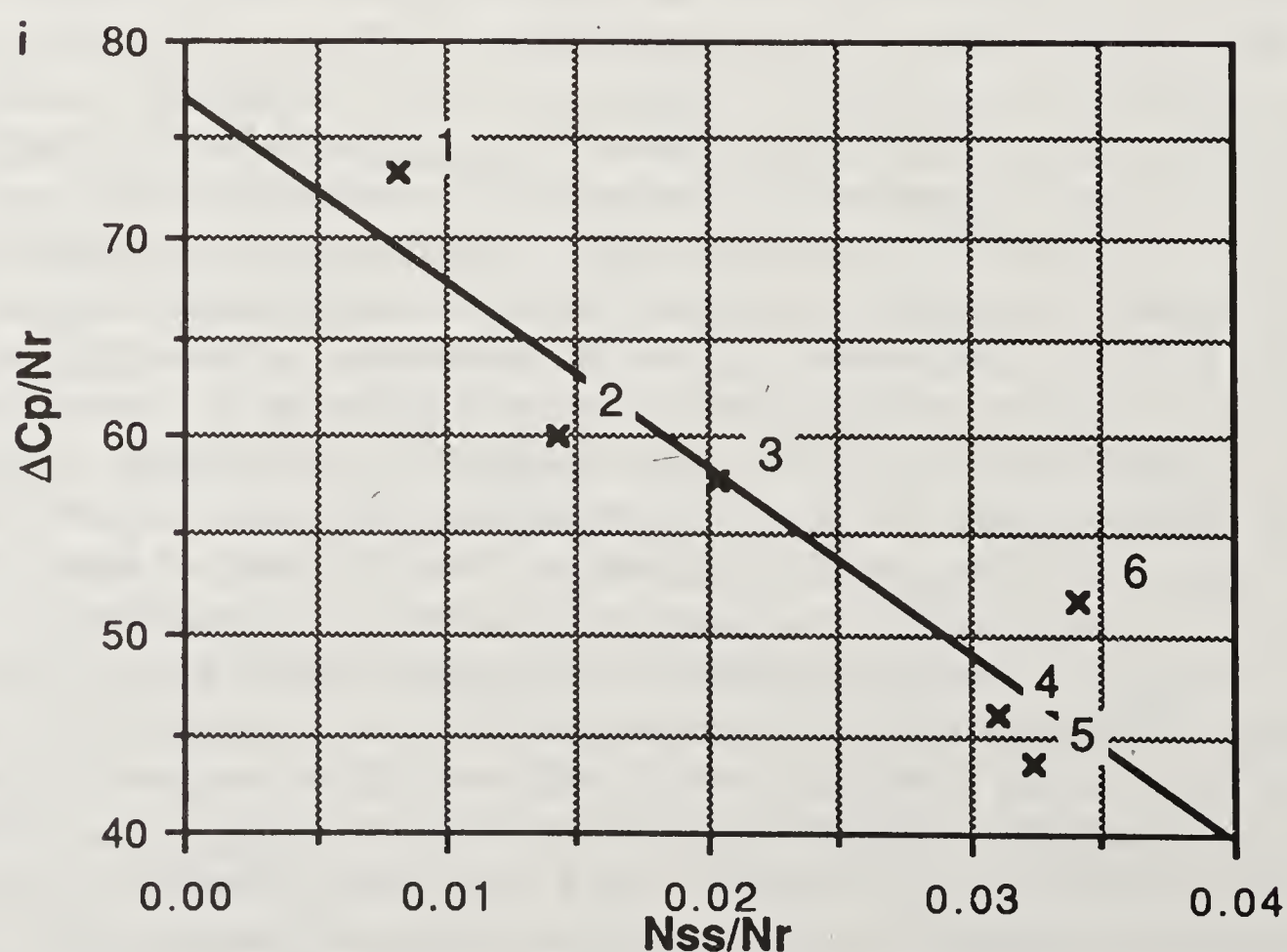


Figure 7. (Caption on facing page.)

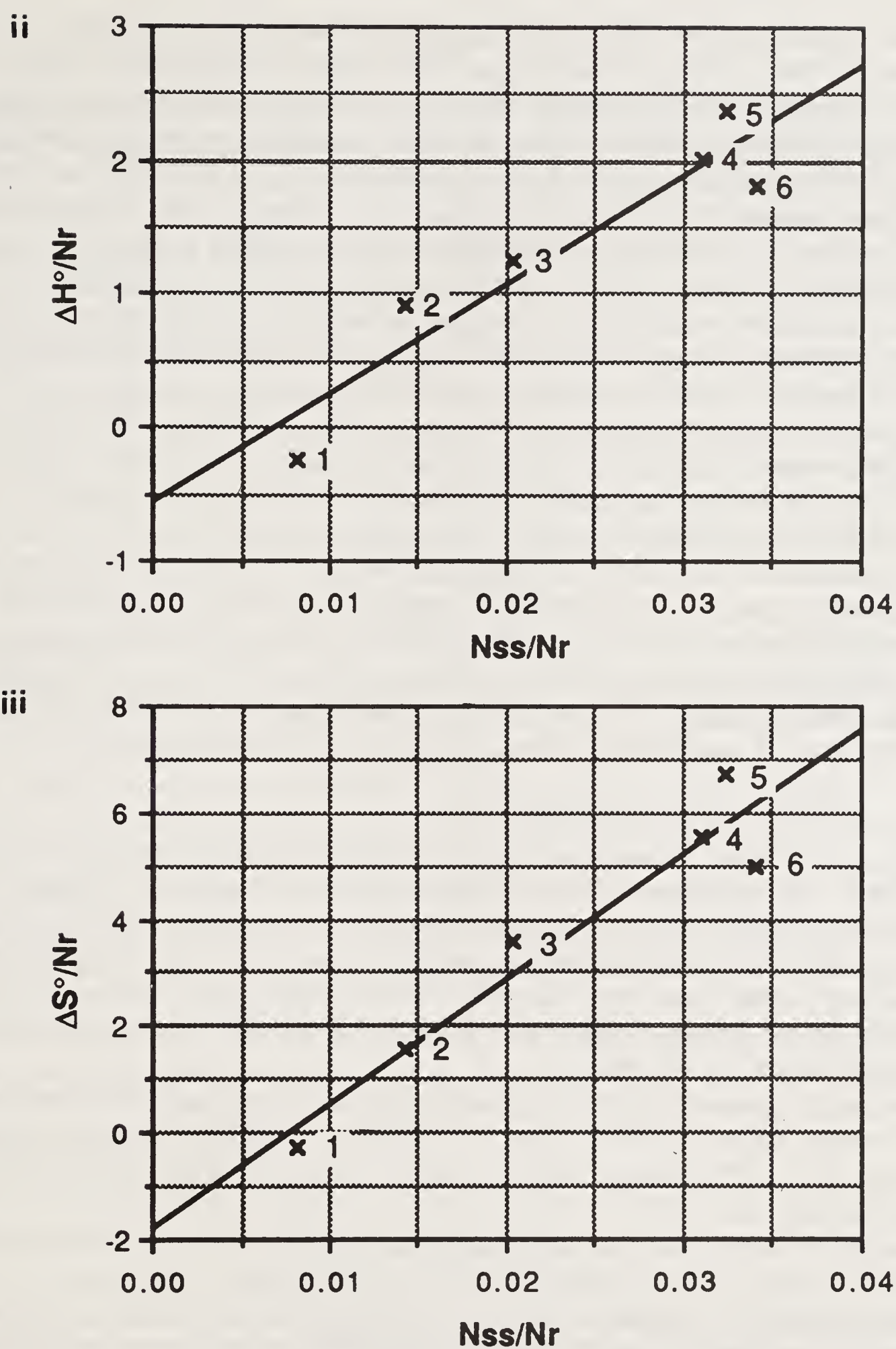


Figure 7. Effects of disulphide bonds on the thermodynamics of protein folding.

Similarly we plot $\Delta H/n_r$ against n_{ss}/n_r (figure 7(ii)) to obtain:

$$\Delta H/n_r = -571 - 82\,500n_{ss}/n_r \text{ (Jmol}^{-1}\text{)}, (R = 0.94). \quad (7)$$

$$\Delta H = -571n_r - 82\,500n_{ss} \text{ (Jmol}^{-1}\text{)}. \quad (8)$$

Plotting $\Delta S/n_r$ against n_{ss}/n_r (figure 7(iii)) gives:

$$\Delta S/n_r = -1.78 + 234n_{ss}/n_r \text{ (JK}^{-1}\text{mol}^{-1}\text{)}, (R = 0.96), \quad (9)$$

$$\Delta S = -1.78n_r + 234n_{ss} \text{ (JK}^{-1}\text{mol}^{-1}\text{)}. \quad (10)$$

These equations show that ΔC_p , ΔH and ΔS for protein unfolding are simple functions of the number of residues and the number of disulphide bonds for the proteins considered (those without metals or cofactors). Since $\Delta C_p/n_r$, $\Delta H/n_r$ and $\Delta S/n_r$ are all proportional to n_{ss}/n_r , all these quantities are proportional to each other (65). The equations show how disulphides stabilise proteins; they disfavour folding entropically but favour it enthalpically to a larger extent. These results may be understood by considering the addition of a disulphide bridge to an unfolded protein. When two regions of the coil are brought together by a disulphide cross-link, the water accessible surface area is decreased, as seen above. Fewer non-polar groups are now exposed to water in the unfolded state, so the magnitude of the hydrophobic effect for protein unfolding decreases. At room temperature, the size of the hydrophobic effect is proportional to ΔC_p and to $-T\Delta S$ (see above). The addition of a disulphide means that ΔC_p decreases and ΔS increases. The change in ΔH cannot be explained by the hydrophobic effect since the enthalpic component of the hydrophobic effect is approximately zero at room temperature (see above). Instead we believe that when a so-called random coil cannot be truly random because it is constrained by one or more disulphide bonds, the hydrogen-bond network of the system (including the water) is less favourable than in an unconstrained system. Thus ΔH of unfolding is increased by the presence of disulphide bonds. It is also likely that disulphides perturb the energetics of the folded state (50). Combining the equations for ΔH and ΔS with $\Delta G = \Delta H - T\Delta S$ at 298 K, we obtain:

$$\Delta G = -41n_r + 12\,800n_{ss}, \text{ Jmol}^{-1}. \quad (11)$$

Thus, on the basis of the present data set, each disulphide stabilises a protein by 12.8 kJmol^{-1} , in agreement with an average experimental value of 12.6 kJmol^{-1} (66). When successive disulphides are broken the protein becomes less stable by this amount until it unfolds. If all the disulphides are broken the protein is unstable by $41n_r \text{ Jmol}^{-1}$, a remarkably small value which emphasises the marginal nature of protein stability. Unfolding on cleavage of disulphides has been observed for ribonuclease A (66), BPTI (67) and ribonuclease T1 (66).

Since ΔC_p is proportional to ΔA_{np} (see above), and each disulphide decreases ΔC_p by an average value of $936 \text{ JK}^{-1}\text{mol}^{-1}$ (equation 6), it follows that each disulphide decreases ΔA_{np} by an average value of 590 \AA^2 . This is the effect the introduction of each disulphide has on the non-polar surface area of the unfolded state. Stabilisation of proteins by disulphides has previously been ascribed to a decrease in the conformational entropy of the unfolded chain (66, 68–70). The magnitude of this effect depends on the size of the loop which is cross-linked. Though this effect is undoubtedly real, it is outweighed by the effect of the disulphides on the hydrophobic effect and hydrogen-bonding. The average size of the conformational entropy is $-57 \text{ JK}^{-1}\text{mol}^{-1}$ per disulphide for the 6 proteins analysed above. The total average effect of each disulphide on the entropy of unfolding is $234 \text{ JK}^{-1}\text{mol}^{-1}$. We believe that the difference between these two values ($291 \text{ JK}^{-1}\text{mol}^{-1}$) is due to the effect of the disulphide on the hydrophobic effect, via a decrease in the surface area of the unfolded state. Since, at room temperature, ΔG for the hydrophobic effect is proportional to the ΔS and ΔA_{np} , we estimate that each disulphide decreases ΔA_{np} , 690 \AA^2 , in good agreement with the previous estimate of 590 \AA^2 . We stress that the equations and average figures obtained by this analysis apply only to the proteins in our data set, that is, globular proteins with disulphide bonds and no coordinated metals or cofactors.

Conclusion

In this article we have considered the problem of molecular recognition, focussing specifically on peptides and proteins. In the introduction we discussed the general importance of the subject, illustrated with examples taken from naturally occurring systems. In order to investigate the forces which operate when molecules bind to one another it is necessary to have a well-understood receptor/substrate complex. Our chosen model is the vancomycin group of antibiotics which bind to simple peptides, analogues of bacterial cell walls. The structure elucidation of vancomycin and ristocetin were described, followed by a discussion of their mode of action in qualitative terms. If the mode of action is to be understood more precisely it is necessary to make quantitative measurements on each of the forces which contribute to the entropy and enthalpy of binding. The factors responsible for the thermodynamics of peptide-peptide recognition include translational and rotational entropy, complementarity between receptor and substrate, non-covalent bonding including electrostatic, van der Waals and hydrogen bonding, the hydrophobic effect and internal rotations about single bonds. One method of understanding each of these is to make small changes to the peptide which binds to ristocetin and measure differences in binding entropy and enthalpy. For example, a comparison of the binding of Ac-D-Ala and Ac-Gly-D-Ala can show the contribution of an amide-amide hydrogen-bond. A second method of measuring the forces between peptides is to discover statistical correlations between thermodynamic measurements and the groups which contribute to the thermodynamics. This method has led to the determination of the contribution of disulphide bonds to protein stability. It was shown that disulphides surprisingly stabilise proteins enthalpically but destabilise them entropically. This result was interpreted as the disulphide bond decreasing the non-polar surface area of the unfolded protein. An understanding of the hydrophobic effect has led to estimates of the change in non-polar surface area on protein unfolding. This confirms that cross-links decrease the surface area of unfolded proteins. An analysis of the surface area of myoglobin, which is uncomplicated by cross-links when unfolded, indicates that the unfolded state may consist of mixtures of residues possessing the dihedral angles of α -helices or β -sheets.

Acknowledgements

We wish to thank past and present members of the DHW research group for their contributions to the research. SERC (UK), the Upjohn Company, Kalamazoo and ICI are thanked for financial support.

References

1. Williams, D. H., Stone, M. J., Hauck, P. R. & Rahman, S. K. (1990) *J. Nat. Prod.* **52**, 1189–1208.
2. McCormick, M. H., Stark, W. M., Pittenger, R. C. & McGuire, G. M. (1955–56) *Antibiotics Annu.* 606–611.
3. Tedesco, F., Markham, R., Gurwith, M., Christie, D. & Bartlett, J. G. (1978) *Lancet* 226–228.
4. Marshall, F. J. (1965) *J. Med. Chem.* **8**, 18–22.

5. Johnson, C. R. (1962) PhD Thesis, University of Illinois.
6. Williams, D. H. & Kalman, J. R. (1977) *J. Am. Chem. Soc.* **99**, 2768–2774.
7. Weringa, W. D., Williams D. H., Feeney, J., Brown, J. P. & King, R. W. (1972) *J. Chem. Soc., Perkin Trans 1*, 443–446.
8. Johnson, A. W., Smith, R. M. & Guthrie, R. D. (1972) *J. Chem. Soc., Perkin Trans. 1*, 2153–2159.
9. Smith, K. A., Williams D. H., & Smith, G. A. (1974) *J. Chem. Soc., Perkin Trans. 1*, 2369–2376.
10. Smith, G. A., Smith, K. A. & Williams, D. H. (1975) *J. Chem. Soc., Perkin Trans. 1*, 2108–2115.
11. Sheldrick, G. M., Jones, P. G., Kennard, O., Williams, D. H. & Smith, G. A. (1978) *Nature (London)* **271**, 223–225.
12. Williamson, M. P. & Williams, D. H. (1981) *J. Am. Chem. Soc.* **103**, 6580–6585.
13. Kalman, J. R. & Williams, D. H. (1980) *J. Am. Chem. Soc.* **102**, 897–905.
14. Williams, D. H., Rajananda, V., Bojeson, G. & Williamson, M. P. (1979) *J. Chem. Soc., Chem. Commun.* 906–909.
15. Harris, C. M. & Harris, T. M. (1982) *J. Am. Chem. Soc.* **104**, 363–365.
16. McGahren, W. J., Martin, J. H., Morton, G. O., Hargreaves, R. T., Leese, R. A., Lovell, F. M., Ellestad, G. A., O'Brien, E. & Holker, J. S. E. (1980) *J. Am. Chem. Soc.* **102**, 1671–1684.
17. Ellestad, G. A., Leese, R. A., Morton, G. O., Barbatschi, F., Gore, W. E., McGahren, W. J. & Armitage, I. M. (1981) *J. Am. Chem. Soc.* **103**, 6522–6526.
18. Michel, K. H., Shah, R. M. & Hamil, R. L. (1980) *J. Antibiotics* **33**, 1397–1406.
19. Debono, M., Molloy, R. M., Barnhart, M. & Dorman, D. E. (1980) *J. Antibiotics* **33**, 1407–1416.
20. Hunt, A. J. (1983) *J. Am. Chem. Soc.* **105**, 4463–4468.
21. Hunt, A. H., Debono, M., Merkel, K. E. & Barnhart, M. (1984) *J. Org. Chem.* **49**, 635–640.
22. Hunt, A. H., Elzey, T. K., Merkel, K. E. & Debono, M. (1984) *J. Org. Chem.* **49**, 641–645.
23. Hammond, S. J., Williamson, M. P., Williams, D. H., Boeck, L. D., & Marconi, G. G. (1982) *J. Chem. Soc., Chem. Commun.* 344–346.
24. Hammond, S. J., Williams, D. H. & Nielsen, R. V. (1983) *J. Chem. Soc., Chem. Commun.* 116–117.
25. Perkins, H. R. (1969) *Biochem. J.* **111**, 195–205.
26. Brown, J. P., Feeney, J. & Burgen, A. S. V. (1975) *Mol. Pharmacol.* **11**, 119–125.
27. Kalman, J. R. & Williams, D. H. (1980) *J. Am. Chem. Soc.* **102**, 906–912.
28. Williams, D. H. & Waltho, J. P. (1988) *Biochem. Pharmacol.* **37**, 133–141.
29. Waltho, J. P., Williams, D. H., Stone, D. J. M. & Skelton, N. J. (1988) *J. Am. Chem. Soc.* **110**, 5638–5643.
30. Waltho, J. P. & Williams, D. H. (1989) *J. Am. Chem. Soc.* **111**, 2475–2480.
31. Hildebrand, J. H. & Scott, R. L. (1950) *The Solubility of Nonelectrolytes* (Reinhold, New York).
32. Atkins, P. W. (1982) *Physical Chemistry* (Oxford University Press, Oxford).
33. Page, M. I. & Jencks, W. P. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 1678–1683.
34. Fersht, A. R., Shindler, J. S. & Tsui, W.-C. (1980) *Biochemistry* **19**, 5520–5524.
35. Fersht, A. R. (1981) *Nucleic Acids Res.* **9**, 4627–4637.
36. Fersht, A. R. (1985) *Enzyme Structure and Mechanism* (W. H. Freeman and Company, New York).
37. Kellis, J. T., Nyberg, K. & Fersht, A. R. (1989) *Biochemistry* **28**, 4914–4922.
38. Doig, A. J. & Williams, D. H. (unpublished).
39. Creighton, T. E. (1984) *Proteins* (W. H. Freeman and Company, New York).
40. Kauzmann, W. (1959) *Adv. Protein Chem.* **14**, 1–63.
41. Tanford, C. (1962) *J. Am. Chem. Soc.* **84**, 4240–4247.
42. Tanford, C. (1964) *J. Am. Chem. Soc.* **86**, 2050–2059.
43. Nozaki, Y. & Tanford, C. (1971) *J. Biol. Chem.* **246**, 2211–2217.
44. Nemethy, G. & Scheraga, H. (1962) *J. Phys. Chem.* **66**, 1773–1789.
45. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.

46. Baldwin, R. L. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8069–8072.
47. Privalov, P. L. (1979) *Adv. Protein Chem.* **33**, 167–241.
48. Sturtevant, J. M. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2236–2240.
49. Hawkes, R., Grutter, M. G. & Schellmann, J. A. (1984) *J. Mol. Biol.* **175**, 195–212.
50. Privalov, P. L. & Gill, S. J. (1988) *Adv. Protein Chem.* **39**, 191–234.
51. Hermann, R. B. (1972) *J. Phys. Chem.* **76**, 2754–2759.
52. Reynolds, J. A., Gilbert, D. B. & Tanford, C. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2925–2927.
53. Chothia, C. (1974) *Nature (London)* **248**, 338–339.
54. Privalov, P. L. & Khechinashvili, N. N. (1974) *J. Mol. Biol.* **86**, 665–684.
55. Edsall, J. T. (1935) *J. Am. Chem. Soc.* **57**, 1506–1507.
56. Gill, S. J. & Wadso, I. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2955–2958.
57. Spolar, R. S., Ha, J.-H. & Record, M. T. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8382–8385.
58. Pitzer, K. S. (1953) *Quantum Chemistry* (Constable, London).
59. Smith, P. W., Chang, G. & Still, W. C. (1988) *J. Org. Chem.* **53**, 1587–1590.
60. Rodriguez-Tebar, A., Vasquez, D., Perez Velazquez, J. L., Laynez, J. & Wadso, I. (1986) *J. Antibiotics* **39**, 1578–1583.
61. Mohamadi, F., Richards, N. G. J., Guida, W. C., Liskamp, R., Lipton M., Caufield, C., Chang, G., Hendrickson, T. & Still, W. C. (1990) *J. Comput. Chem.* **11**, 440–467.
62. Creighton, T. E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5082–5086.
63. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283–438.
64. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379–400.
65. Murphy, K. P., Privalov, P. L. & Gill, S. J. (1990) *Science* **247**, 559–561.
66. Pace, C. N., Grimsley, G. R., Thomsun, J. A. & Barnett, B. J. (1988) *J. Biol. Chem.* **263**, 11820–11825.
67. States, D. J., Creighton, T. E., Dobson, C. M. & Karplus, M. (1987) *J. Mol. Biol.* **195**, 731–739.
68. Schellmann, J. A. (1955) *C. R. Trav. Lab. Carlsberg Ser. Chim.* **29**, 230–259.
69. Flory, P. J. (1956) *J. Am. Chem. Soc.* **78**, 5222–5235.
70. Poland, D. C. & Scheraga, H. A. (1965) *Biopolymers* **3**, 379–399.

Molecular Recognition in Aspartic Proteases

H. Mattras¹, M. Bianchi¹, J. A. Fehrentz², R. A. Boigegrain¹,
L. Chiche², M. A. Coletti-Previero¹ and B. Castro³

¹INSERM U 58, rue de Navacelle, 34000 Montpellier, France

²CCIFE, rue de la Cardonille 34094 Montpellier, France

³Sanofi Chimie, 32 rue Marbeuf, 75008 Paris, France

Introduction

Aspartic proteases, one of the four main families of proteases are structurally characterized by the existence of two homologous domains, each contributing to the active site with the invariant Asp–Thr–Gly triad. The general proteolytic activity pH-optimum lies generally below pH 6; hence, they were formerly called “acid proteases”. It has been shown that, contrary to the serine and cysteine proteases, the hydrolytic catalysis involves no covalent enzyme–substrate complex: an active site driven water molecule binds directly to the reactive peptide bond carbonyl, leading to the labile tetrahedral intermediate (1).

Pepstatin, or pepstatin analogs are general inhibitors of aspartic proteases; these compounds contain a hydroxyl group that contributes to the active site binding by the expulsion of the active water molecule.

Aspartic proteases are endopeptidases with very different functions, from digestive degradation (e.g., pepsin) to control of blood pressure (e.g., renin). Recently, viral aspartic proteases have aroused considerable attention, in particular the dimeric HIV-protease, as targets for anti-viral drugs (2).

The close homology of the two domains of aspartic proteases has been interpreted as the result of a gene duplication from an ancestral monomeric gene, coding for a single-domain protein active as a non covalent dimer (3); recently this ancestral feature was discovered in the modern retroviral aspartic proteases (4).

The secondary structure of aspartic proteases is mainly constituted of more or less extended β -sheets connected by atypical loops. Very few helical sections are present. A striking very long β -hairpin structure belonging to the N-terminal domain, the flap covers the active site across the catalytic cleft. It is supposed to play an important role in the catalytic process, as it appears on crystallographic structures in that the flap has to open to allow the substrate to enter the catalytic cleft.

This paper summarises our contributions concerning:

- new fluorogenic substrates, useful as new powerful tools for the study of aspartic proteases including HIV protease.

- the generation of a new aspartic protease constituted by the dimerization of the N-terminal domain of porcine pepsin, by controlled autolysis of the mother protease.

New Fluorogenic Substrates for Aspartic Proteases

The proteolytic activities are commonly detected by the use of peptides of definite sequences, whose breakdown is followed spectrophotometrically (5, 6) or by HPLC separation of products (7). Fluorescent substrates are widely used with exopeptidases (8, 9) and with the endopeptidases possessing esterase and amidase activities (10): they all have the distinct advantage of sensitivity and have become very popular. We afforded recently a distinct contribution to this field, designing fluorogenic substrates that work with endoproteases on their specific peptide substrates; the fluorogenic potential was achieved by adding a C-terminal tryptophan and an N-terminal intramolecular quenching group. The enzymatic cleavage of the peptide results in a proportional increase in fluorescence emission as hydrolysis proceeds. This approach is of general application to a number of endopeptidases (e.g., collagenase (11)) by adapting the peptide sequence to match the target enzyme specificity.

The choice of the heptapeptide DNP-Ser 1-Gln 2-Asn 3-Tyr 4-Phe 5-Val 6-(D)Trp 7-OH 1 as fluorogenic substrate for pepsin was based on the known cleavage

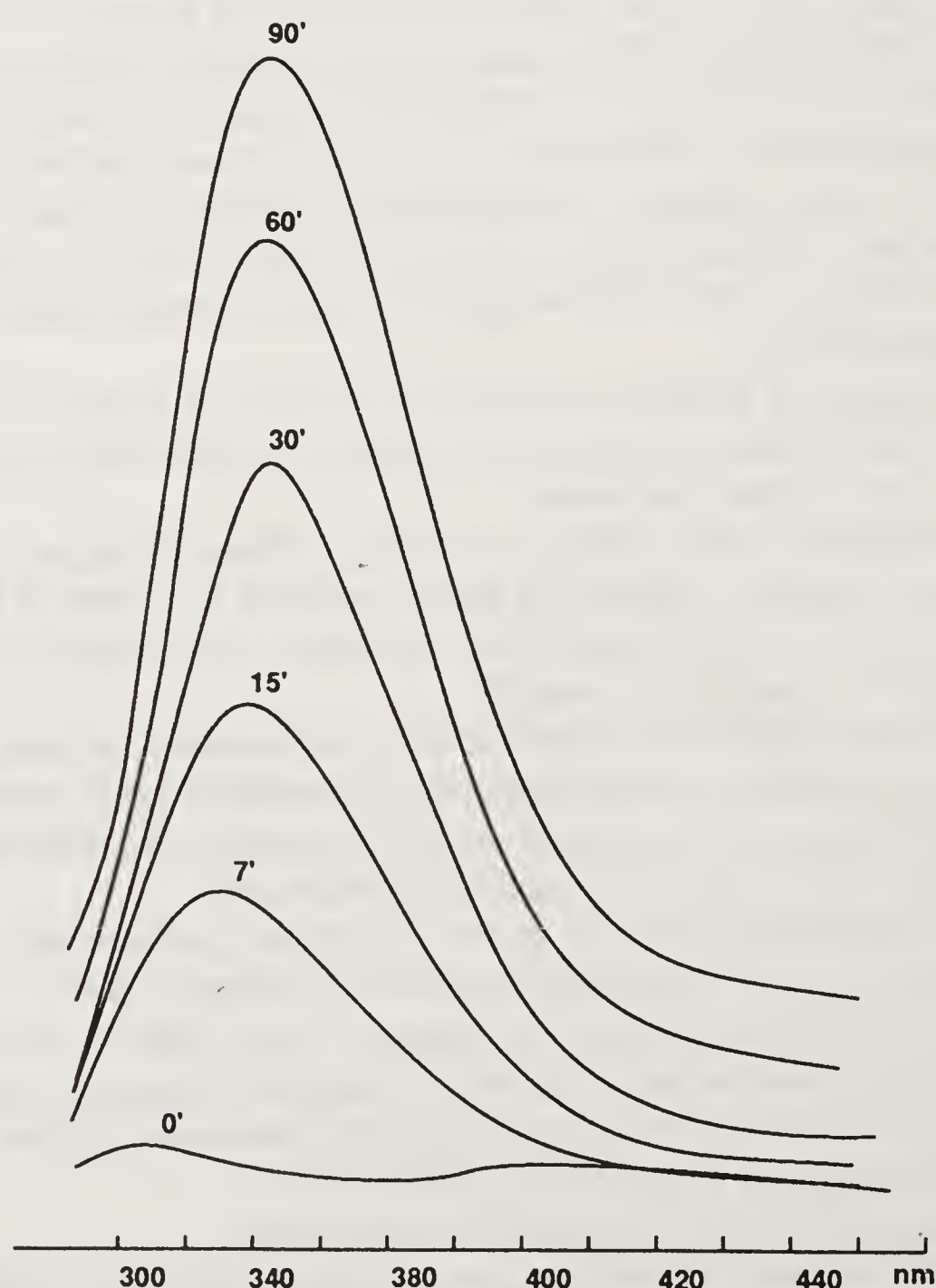


Figure 1. Fluorescence emission spectrum of DNP-Ser-Glu-Asn-Tyr-Phe-Val-(D)Trp-OH (10^{-5} M) before (0) and after pepsin (10^{-6} M) addition (7, 15, 30, 60 and 90 min), pH 4, 25°C. Excitation wavelength was 280 nm.

site of other pepsin substrates. The tryptophan was in its D form to prevent undesired cleavages by other proteases (e.g., carboxypeptidases) eventually present. The DNP group was incorporated to quench the tryptophan fluorescence intramolecularly.

The peptide is a substrate for pepsin, split into DNP-Ser-Gln-Asn-Tyr-Phe-OH and the fluorescent H-Val-(D)Trp-OH. Figure 1 shows the fluorescence emission, recorded at various times after the enzyme was added to the reaction mixture. The starting peptide has no emission spectrum, while a linear increase of fluorescence can be noticed as a function of time (or as a function of enzyme concentration) during the digestion; the maximum is at 346 nm, the characteristic wavelength of tryptophan fluorescence. Thus, the hydrolytic reaction can be monitored at 340 nm and the kinetic parameters evaluated from a single measurement.

This increase can only be attributed to the cleavage of one or more peptide bonds within the peptide chain, which results in the formation of a free NH_2 peptide (no intramolecular quenching group), bearing a tryptophan in C-terminal position. The cleavage of the peptide was confirmed by HPLC separation of the products and their identification was achieved by amino acid analysis of the separated peaks, after acid hydrolysis. The enzymatic reaction was carried out at different pH values and the optimum for pepsin was found to be at pH 4(12). While pepsin optimum pH is around 2 with high molecular weight substrates (e.g., hemoglobin), other synthetic oligopeptides have shown a similar shift in optimum pH, when submitted to pepsin action. The same substrate is also the substrate of chymotrypsin, subtilisin and thermolysin with different cleavage points, as indicated in table 1.

This approach provides several distinct advantages over other substrates for aspartic proteases. First the entire time course of the reaction can be followed with ease and adaptability; second the peptide sequence can be adapted to whatever endoproteolytic activity one is dealing with, provided the peptide is not too long (we tested up to a nonapeptide); third, the sensitivity of fluorescence allows the detection of enzymes in small concentrations (we could detect enzyme levels as low as 10^{-11} M, within reasonable times), an essential parameter when dealing with valuable material.

A New Product Derived Pepsin Inhibitor

In the preceding investigation, we questioned the need of a spacer between the recognition sequence and the fluorescent probe. For that purpose, we tested two other

Table 1. Kinetic constants for DNP-Ser-Glu-Asn-Tyr-Phe-Val-(D)Trp-OH hydrolysis by different proteases.

Enzyme (pH)	km μM	k_{cat} min^{-1}	k_{cat}/km $\mu\text{M}^{-1} \text{min}^{-1}$	Cleavage Products
Pepsin (4)	12.5	31	2.5	DNP-Ser-Glu-Asn-Tyr-Phe-OH + H-Val-Trp-OH
Chymotrypsin (8)	14.5	0.8	0.05	DNP-Ser-Glu-Asn-Tyr-OH + H-Phe-Val-Trp-OH
Subtilisin (8)	33	25	0.7	n.d.
Thermolysin (8)	100	200	2	DNP-Ser-Glu-Asn-OH H-Tyr-Phe-Val-Trp-OH

peptides where one and two glycine residues were intercalated between Val and Trp: DNP-Ser 1-Gln 2-Asn 3-Tyr 4-Phe 5-Val 6-Gly 7-(D)Trp 8-OH **2**, and DNP-Ser 1-Gln 2-Asn 3-Tyr 4-Phe 5-Val 6-Gly 7-Gly 8-(D)Trp 9-OH **3**.

Very strikingly, if **3** behaved as a good substrate for pepsin, the hydrolysis of **2** was rapidly blocked at about 1 to 2% (related to enzyme concentration) from completion. This phenomenon can be interpreted by the inhibiting properties of the C-terminal fragment. Indeed, the synthetic tripeptide H-Val-Gly-(D)Trp-OH has been found to be an inhibitor ($K_i = 7.5 \times 10^{-6}$ M using the classical yellow peptide substrate Leu-Ser-Phe(NO₂)-Nle-Ala-Leu-OMe) of pepsin, whereas the tetrapeptide H-Val-Gly-Gly-(D)Trp-OH has no inhibiting activity. Conformational studies are currently made to interpret this phenomenon.

A New Fluorogenic Substrate for HIV Protease

The heptapeptide with Pro in position 5 was designed to provide high affinity for HIV protease. The sequence Ser-Gln-Asn-Tyr-Pro represents the p17-p24 function in the HIV-1 gag polypeptide (13, 14) and cleavage of the peptide bond takes place between the aromatic amino acid residue and the Pro residue, as demonstrated by HPLC point assays (15) and continuous spectrophotometric assay (16). There is a general trend for hydrophobic residues in P2' and P3' sites, so Val and Trp were used and the peptide DNP-Ser-Gln-Asn-Tyr-Pro-Val-(D)Trp-OH **4** contains the consensus sequence to be a fluorogenic substrate for HIV-protease. Preliminary assays confirmed this hypothesis (17). The single amino acid interchange yields, as expected, a peptide completely inert to pepsin action.

N-terminal Domain of Pepsin as a Model for Retroviral Dimeric Aspartic Protease

Pepsin and in general non-viral aspartic proteases are composed of two domains, each contributing to the formation of the active site. These enzymes are thought to have evolved from an ancestral protein by gene duplication and subsequent gene fusion. The highly conserved triads of amino acids, Asp-Thr-Gly, located in each topologically similar lobe, suggested that, prior to gene duplication, the primordial enzymes may have acted as dimers. The recently presented evidence that retroviruses encode half an aspartic protease, active as a dimer (18, 19, 20), strengthened this hypothesis.

Studies to determine to what extent the entire polypeptide chain of the aspartic-proteases is needed to achieve activity were reported on pepsin and on renin. The experiments to shorten the pepsin molecule were achieved by autodigestion (21, 22) in an attempt to isolate from the enzyme the smallest possible fragment retaining activity, the "active site". Smaller renin fragments, still active, were isolated from juxtaglomerular cell tumor and became apparent only under strong dissociating conditions. Their presence was attributed to limited proteolytic attack on the peptide connecting the two parts of the protein (23).

This study was undertaken to learn if limited proteolysis of pepsin might result in the formation of a single domain acting as the viral aspartic protease. A 136 amino acid

fragment from the N-terminal domain of porcine pepsin was obtained by controlled autolytic degradation. It has a strong tendency to self-assembly, resulting in an active dimeric form with a specificity toward substrates and inhibitors similar to the parent enzyme.

Formation and Purification of a New Protease

Porcine pepsin (3×10^{-4} M), dialysed against twice the volume of buffer (pH 4) under controlled conditions of temperature and pH (figure 2), undergoes spontaneous degradation; dialysable autolytic fragments, which show both an absorption at 280 nm and proteolytic activity, are released. In order to obtain a reasonable yield of active material a number of parameters such as the pH, the temperature, the concentration, the volume of the sample and the time, required optimisation.

The pH value was determined experimentally to yield an autolytic reaction sufficiently slow to avoid explosive degradation and sufficiently rapid to achieve reasonable yields. The autolysis was started at 4°C, a temperature which allowed a controlled cleavage only at the most exposed polypeptide segments and was then elevated to ensure a better overall yield of the active fragment. If the autolytic cleavage was performed at 4°C the yield was low, if the temperature was enhanced to 25°C from the start no activity was obtained in the dialysate, possibly as a result of unspecific degradation. Times longer than 48 h were unsuitable since a loss in yield was always observed. The concentration of pepsin (10 mg/ml) is the best compromise since it must

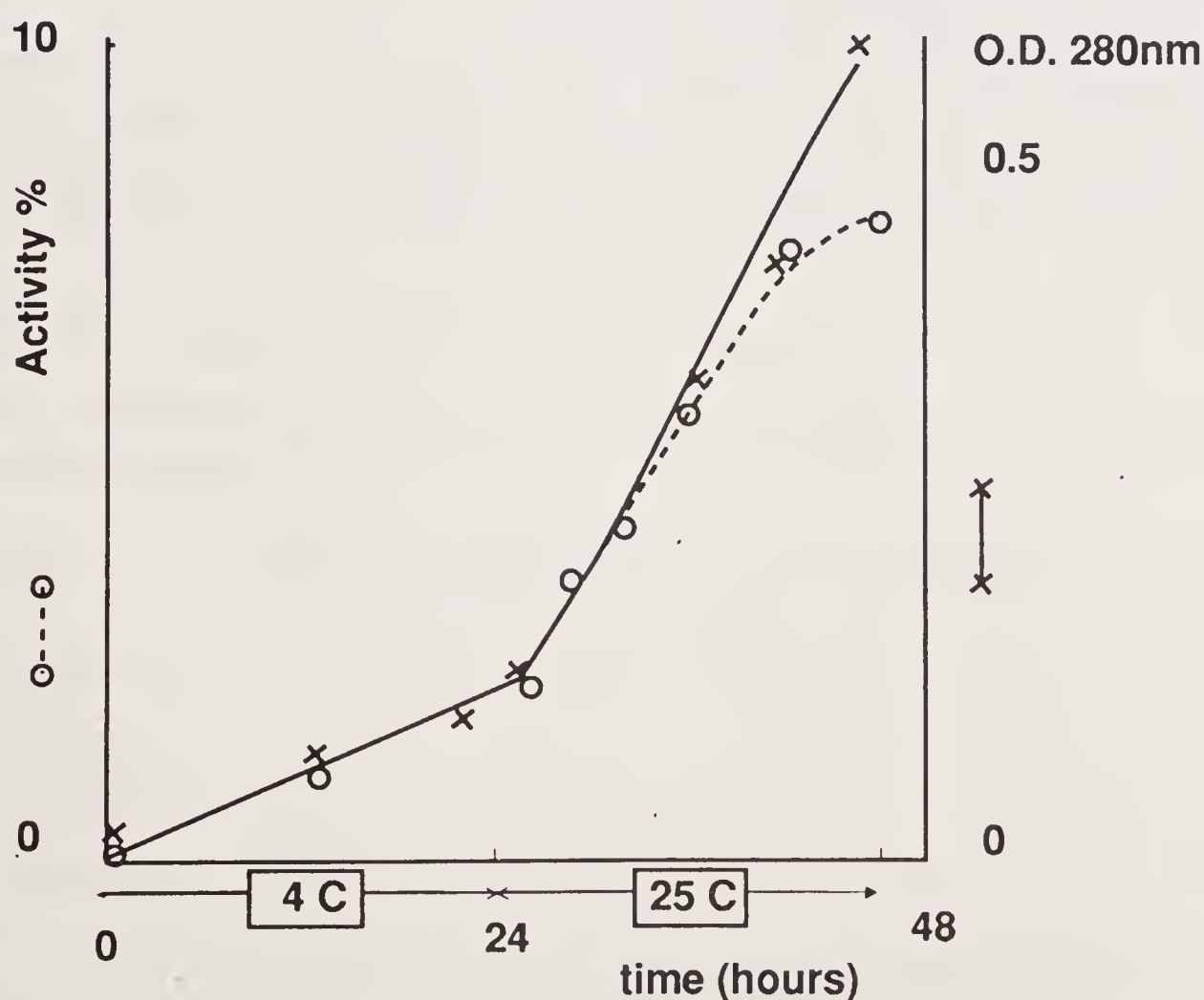


Figure 2. Autolysis of porcine pepsin. The absorbance (—) and proteolytic activity (---) of the dialysate was measured on haemoglobin or Leu-Ser-Phe(NO₂)-Nle-Ala-OMe as a function of time.

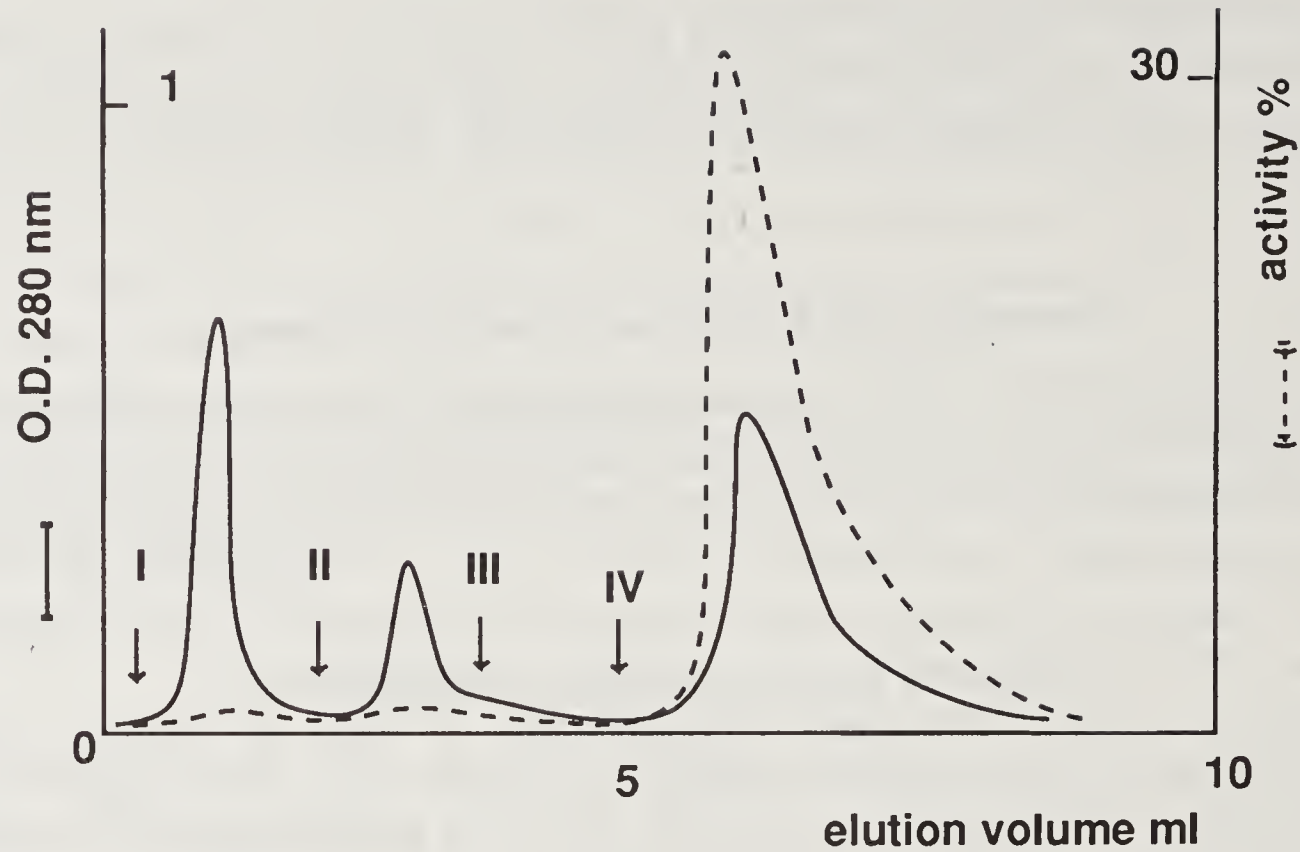


Figure 3. Alumina pseudo-affinity chromatography (29) of the dialysate; the absorbance (—) and proteolytic activity (----). The column was sequentially eluted with: I, 0.1 M acetate buffer, pH 4.2; II, 0.1 M acetate buffer, NaCl 0.5 M pH 4.2; III, water; IV, 0.1 M phosphoric acid, pH 4.

act as an enzyme and as a substrate and the choice of the volumes was made to insure the best possible drain of active material from the bag. Finally the glassware must be of good quality to avoid losses by adhesion.

The first event to take place should be a limited proteolytic attack on the connecting peptide between the two domains, an assumption verified experimentally in a number of other enzymes containing substructural domains (24, 25, 26, 27). On the basis of the dialysis bag pore size, the theoretical MW of the molecules able to leave it should not exceed 14,000/16,000. Since the MW of the N-terminal domain of porcine pepsin is about 16,000 and that of the C-terminal one about 18,000, the resulting fragments could escape through the dialysis bag, if partly degraded.

Porcine pepsin is a phosphoprotein (28), with a phosphorylated serine at amino acid 68 in the first domain. The chromatography on alumina, pH 4.2, of the dialysed material showed that practically all the activity was retained on aluminium oxide and could be displaced by inorganic phosphate (figure 3). After desalting, the product eluted

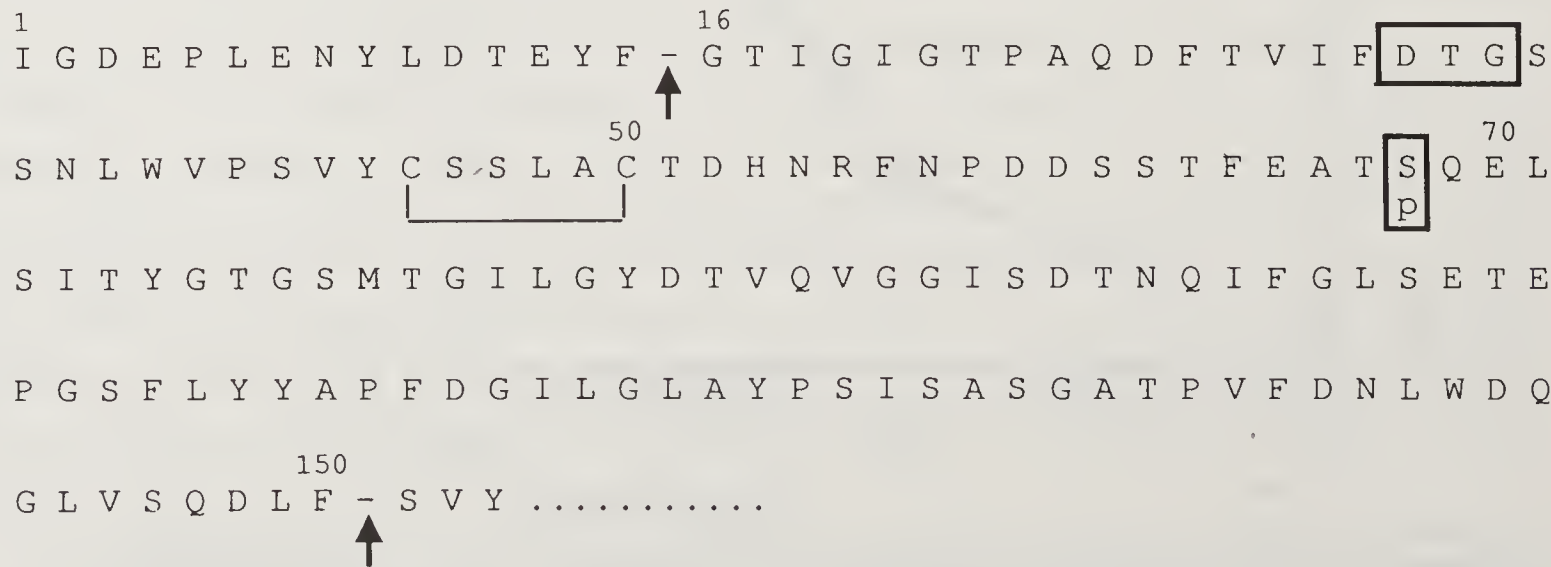


Figure 4. First domain of porcine pepsin. Boxed, the active triad and the phosphorylated Ser 68.

from the molecular sieve chromatography as a symmetrical peak in the range of 30,000 MW and the experimental amino acid composition compared well with that calculated from the known amino acid sequence of the N-terminal domain of the parent porcine pepsin (figure 4). The absence of basic amino acids (present only in the C-terminal domain of porcine pepsin) ruled out the presence of the second domain, which is either more slowly or not dialysable, more susceptible to a disruptive fragmentation, or is frontally eluted during the alumina chromatography. Chemical and enzymatic analysis of the N-terminal and C-terminal allowed the identification of the active fragment as the 16–151 fragment from the parent pepsin (29). The actual MW of this fragment being about 15,000 with only one Asp–Thr–Gly triad per molecule, the only possibility to account for the experimental MW of 30,000 together with the need of two aspartates for enzymatic activity is a dimerisation of the fragment. A summary of the isolation and identification of the active fragment is presented in table 2.

However the affinity chromatography of very dilute samples can be difficult: the sample must be either concentrated, which is sometimes improper, or eluted with much longer times to get a sufficient contact and achieve a reasonable interaction yield. We took advantage of the lack of back pressure of CPC alumina to set up a recycling apparatus (figure 5) consisting of a circulating pump, connected with a packed-bed reactor containing the support; the outlet downward flow is recirculated to the reservoir. The temperature of the system is controlled by a thermostated bath.

In the case of the active fragments of pepsin the support was underivatized alumina and the yields were dramatically increased when compared to the ones previously reported (30). This was attributed to the fact that the lyophilisation step could be avoided and that the loss due to the dilution was lowered by the numerous recyclings of the solution through the reactor. The same recycling technique can be used to perform the affinity chromatography of other ligands with suitably derivatized alumina (31, 32, 33) and with the same advantages, which are – the concentration of the sample during purification, the simplicity and the versatility of its use and the good to very good yields it affords. The dimer model (see figure 9 below) shows that the phosphoserine 68 is very accessible and conveniently located to achieve an excellent interaction with the solid alumina.

Table 2. Purification of the active fragment from porcine pepsin.

Preparation	Total mg protein	Hb activity (Absorbance/min)	SA	Peptide activity (Absorbance/min)	SA	Yield %
Dialysate, pH4	10	2.45	0.24	100	0.1	—
Alumina						
Pseudo-affinity	1.5	2.1	1.4	85	56	85
Chromatography						
Molecular sieve	0.03	0.86	28	25	833	25
Chromatography						

In the isolation of the autodigestion products from 100 mg of porcine pepsin, the discrepancy between the specific activities (SA) measured on Hb and on the synthetic peptide are due to the preference shown by the active fragment for small MW substrates. The amino acid analysis was in excellent agreement with the composition of fragment 16–151 from porcine pepsin (see figure 3).

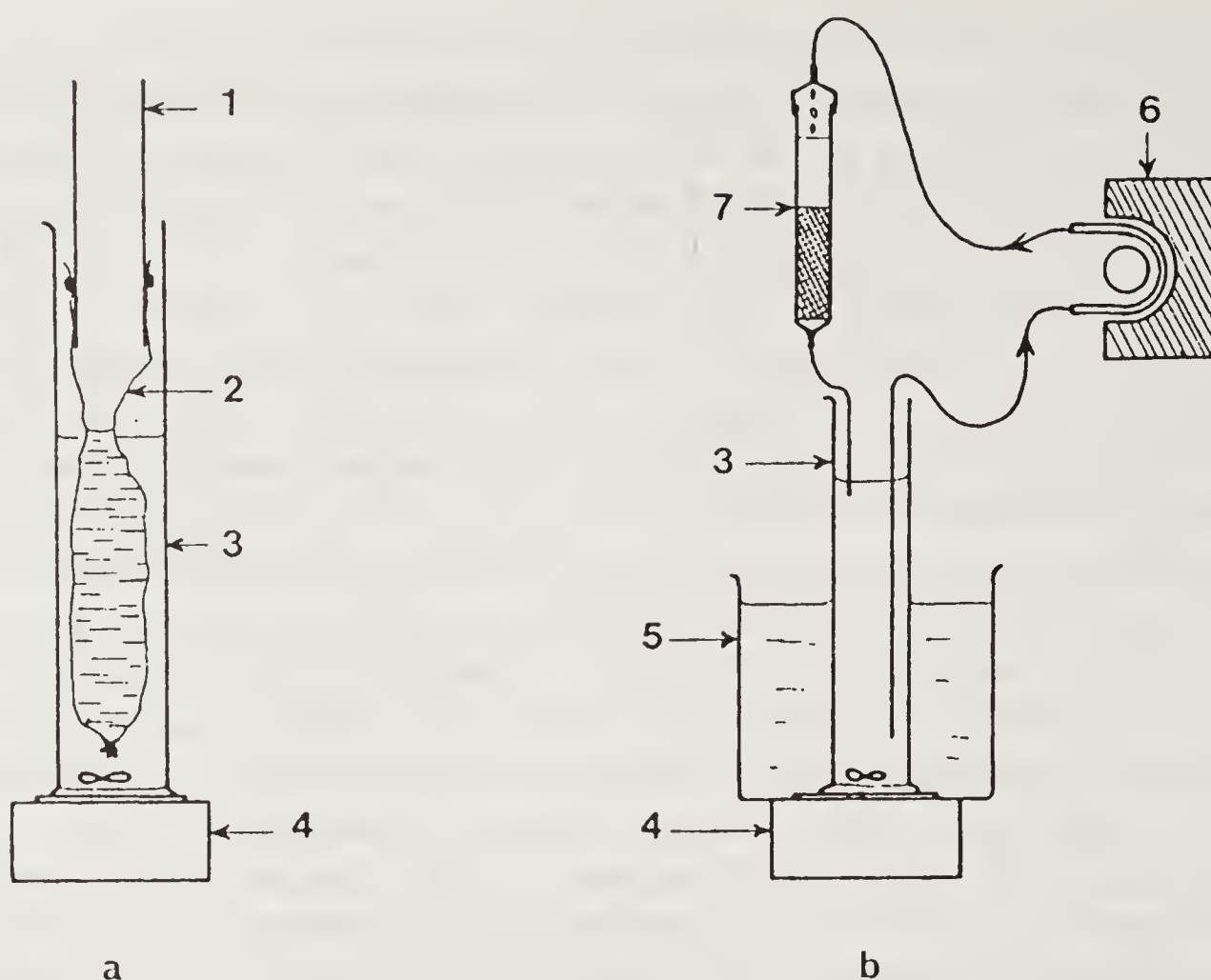


Figure 5. Dialysis (a) and Apparatus (b) to perform affinity chromatography of diluted samples. 1 = glass tube; 2 = Visking dialysis bag; 3 = cylindrical container; 4 = magnetic stirrer; 5 = thermostated bath; 6 = circulating pump; 7 = packed-bed reactor.

Enzymatic Properties

The digestion of hemoglobin and of the synthetic substrate Leu-Ser-Phe(NO₂)-Nle-Ala-Leu-OMe by the active fragment was studied and the results compared with the proteolysis by the parent enzyme. It soon appeared that hemoglobin, which is an excellent substrate for pepsin, is a very poor one for the autolytic fragment, the k_{cat} being lower than one order of magnitude. The dimer activity, estimated to be tenfold less (30) was probably still overestimated, due to the presence in the crude commercial hemoglobin used as substrate of shorter fragments of the protein. The synthetic peptide, by contrast, is a much better substrate ($k_{\text{cat}}/\text{km porcine pepsin} = 5 \text{ mM}^{-1} \text{ s}^{-1}$; $k_{\text{cat}}/\text{km autolytic fragment} = 240 \text{ mM}^{-1} \text{ s}^{-1}$). This difference is the consequence of the low temperature dependence of the dimer activity (see below). The low activity against the protein substrate could be interpreted by the fact that the insertion of a more rigid protein substrate within the catalytic cleft requires an adaptative deformation of the enzyme that the dialysable fragment is unable to afford without dissociation. The pH-dependence of the enzymatic catalysis is significantly the same (optimum pH: 2.2 on hemoglobin, 4.5 on synthetic substrate). Pepstatin inhibited the digestion of the synthetic substrate with a K_i of 10^{-11} M , which further demonstrates that the dimeric fragment behaves as an aspartic protease.

One of the more striking dissimilarities between pepsin and its autolytic fragment was the better resistance of the dimer to basic pH and to autolysis. If the activity of both enzymes (at pH 4.7 on the synthetic substrate Leu-Ser-Phe(NO₂)-Nle-Ala-Leu-OMe) is studied as a function of their resistance to pH denaturation, a slightly alkaline pH, which denatures pepsin is without effect on the dimer (figure 6); best results are obtained in the presence of 1% Triton X-100 to avoid undue precipitations and/or

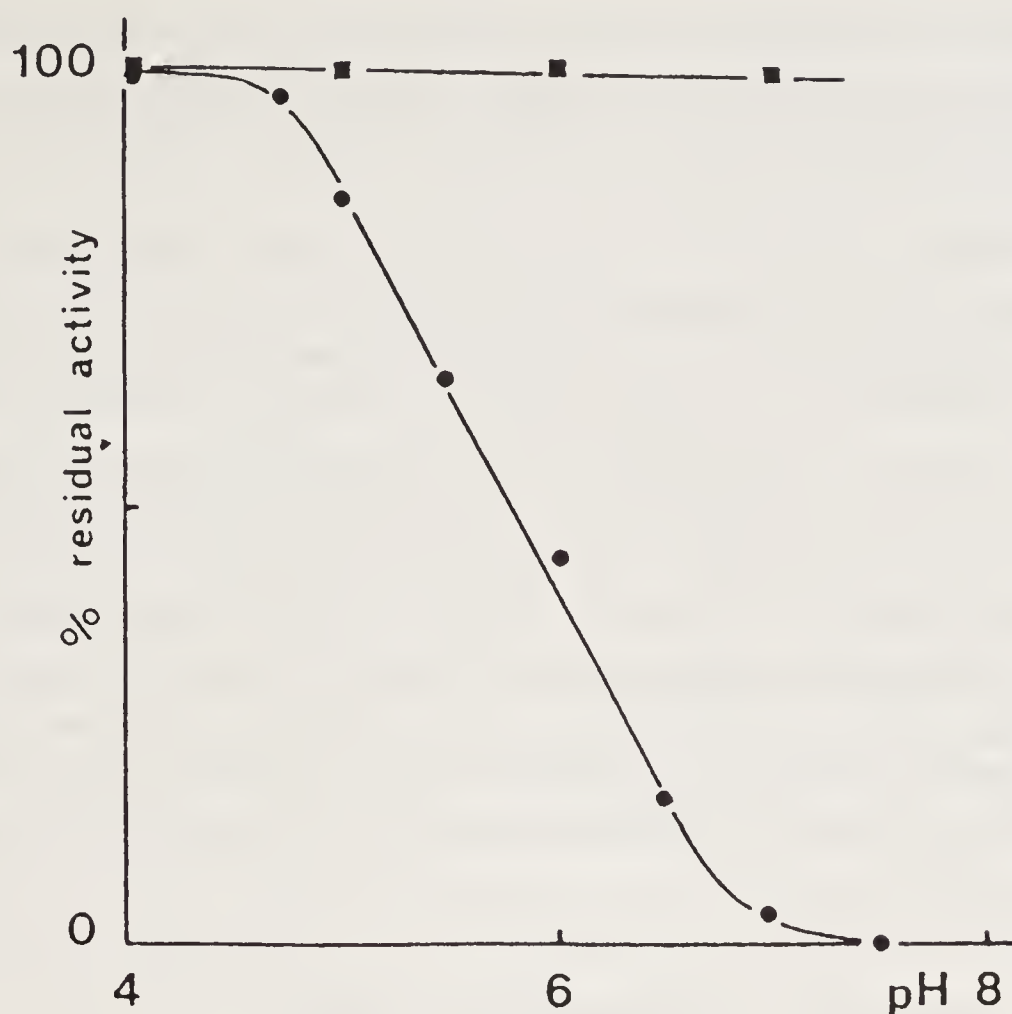


Figure 6. Effect of pH on the activity of pepsin (●) and of the dimer (■). Concentrated samples were allowed to stay at pH for 10 min and were then diluted to pH 4.7 for activity. All solutions 1% in Triton X-100.

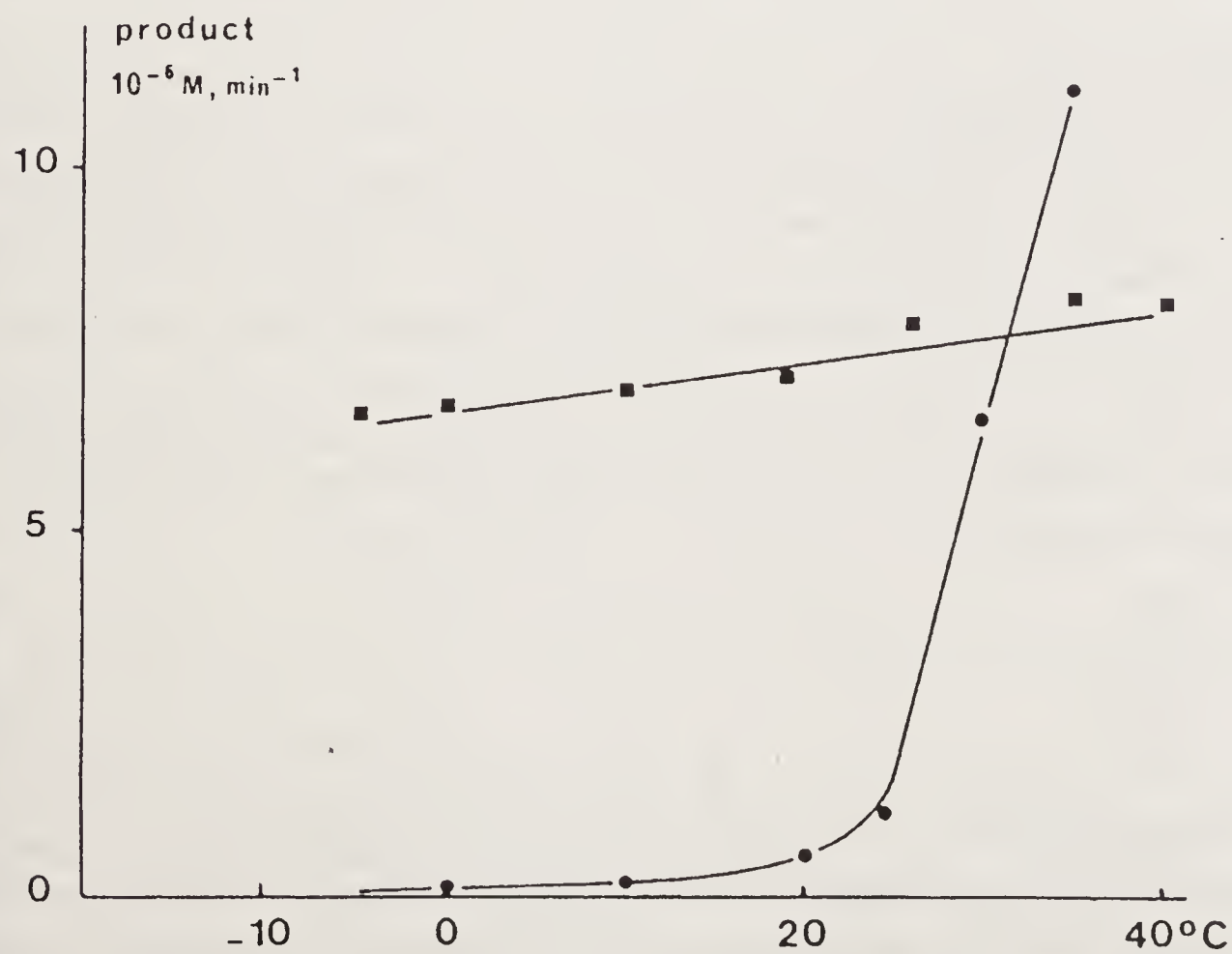


Figure 7. Activity of pepsin (●) and of the dimer (■) at different temperatures. all solutions 1% in Triton X-100.

adsorptions on the glass. The rate of hydrolysis of the same synthetic substrate by the two aspartic proteases, measured at increasing temperatures is strikingly different (figure 7). Surprisingly enough the dimer is still fairly active at -5°C on the synthetic substrate.

A number of conclusions can be drawn from these results. First, even if a small leakage of pepsin from the dialysis bag cannot be ruled out since pepsin would coelute with the dimer from alumina, it seems undetectable at the end point of the purification and certainly is not responsible for the proteolysis. Second, the dimer, although generated by pepsin, is on the overall a different aspartic protease, more stable to pH, able to perform at subzero temperatures and much more resistant to autodigestion.

It can be asked at this point why pepsin autodegrades to complete or nearly complete inactivity. Since an autolytic dimeric fragment more resistant to autolysis is formed, a steady state should appear and from that point on the autodegradation rate should be significantly lower. The answer could be that, at the optimal conditions of pepsin digestion (pH 2–3, temperature higher than 4°C) the yield of the fragment is so low and its half-life so short that it will not be able to accumulate and its presence would go unnoticed.

Molecular Modelling of the Protease

As no all-atom coordinates are available for the porcine pepsin, the three-dimensional structure of porcine pepsinogen (34) has been used as a starting point for the modelling. In order to make sure that no significant deviation could be ascribed to the presence of the prosegment in the pepsinogen, the 16–151 fragment was first compared to rhizopuspepsin (35). It was found that the fragment fits perfectly onto the corresponding rhizopuspepsin segment with an rms deviation of 0.6 \AA for 105 C- α pairs superimposed (out of 136) and can be used further without modification. The pepsin dimer was then built by superimposing the isolated 16–151 segment successively on the two chains of HIV-1 protease (36), since retroviral proteases have been shown to be dimeric species similar in structure and in properties to the monomeric cellular aspartic proteases (37, 38). For the 36 C- α pairs fitted the rms deviation was 0.79 \AA . The N-terminus of the segment is obviously too short to be involved in a four-stranded β -sheet similar to the one found at the interface of the dimer in the retroviral proteases. On the other hand the C-terminal helix (denoted h' in HIV-1 protease) is longer in the dimer (three turns instead of two) and shifted away from the interface. The residue 144 at the end on the helix is then located 10 \AA away from residue 93 at the end of the helix in HIV-1. The alignment between the two C-terminals was therefore based mainly on geometrical considerations:

	88																		98
HIV-1	N	L	L	T	Q	I	—	—	—	—	—	—	—	G	C	T	L	N	
	135																		151
Pepsin	P	V	F	D	N	L	W	D	Q	G	L	V	S	Q	D	L	F		

Residues 147 through 151 were manually superposed to the 94–98 HIV-1 segment and then a regularization was carried out on the 135–151 segment keeping the ends fixed and preserving the helical secondary structure between 135 and 144 via angle constraints using FROD (39). The dimerization of the pepsin fragment led to an overlap of the two flap tips which were then modelled on the conformation observed in

native uncomplexed HIV-1 protease. Finally AMBER (40) molecular mechanics minimization (100 steepest descent cycles with constrained main chain followed by 100 steepest descent plus conjugate gradient cycles without constraints) was done to remove the main residual bad contacts: a view of the dimer model from porcine pepsin is shown in figure 8.

The dimer interface in this model is largely modified when compared to the one in the retroviral proteases: only two of the four strands are still present in the β -sheet but new favorable hydrophobic interactions appear between residues Phe 111 and Phe 117 of chain 1 and Tyr 125 and Ile 128 of chain 2 (and of course residues 111 and 117 of chain 2 with residues 125 and 128 of chain 1). These hydrophobic contacts, displayed in figure 9 do not exist in the retroviral proteases, which shows a deletion in these positions. The model obtained was checked against the solvation free energy of folding (41) that was recently proposed as a useful criterion to evaluate protein models (42). The dimer shows a value rather close to globular proteins and thus appears reasonable on the basis of the hydrophobic effect (calculated deviation from the predicted value: + 5.3%). The dimer value is also slightly better than for the isolated monomer (calculated deviation from the predicted value: + 8.6%): this result is consistent with the hydrophobic interface described above and shows that the dimerization process of the 16–151 fragment is probably favored on energetic grounds.

Monomer/Dimer Equilibrium

The active dimer was submitted to the action of the temperature and the decrease in activity was recorded together with the amount of monomer formed (M_r 15,000, elution volume from the G 75 column = 63 ml). Figure 10 shows the effect of increasing temperatures on the activity, measured at room temperature on the synthetic peptide.

It is noteworthy that no inactivated dimer could be detected by molecular sieve chromatography and the loss in activity could be attributed in each case to the formation of an inactive monomeric form. When the dimer in solution was kept at 4°C the decay of activity and the monomer formation were very slow (50% in almost a month, figure 10). No activity retrieval could be obtained from temperature inactivated samples by simple return to room temperature or to 4°C. This tendency to lose both the activity and the dimeric conformation must be taken into account and dictates the use of freshly prepared active dimer. This is achieved by keeping the fragment at 4°C on the alumina support, where there is no indication of loss in activity with time, and by eluting with phosphate buffer (30) just before use.

Dissociation by chaotropic reagents was also experienced. Some of the added chemicals were surprisingly ineffective towards the dimer, table 3 summarizes the results. Treatment of the dimer solution for short times (10 to 15 min) with salts and with urea (up to 3 M) did not affect the activity: dioxane induces a loss of enzyme, possibly due to precipitation. The non ionic detergent Triton X-100 somewhat stabilizes the active molecule.

Only guanidine hydrochloride (GnCl) achieves the dissociation of the dimer and the loss in activity (hydrolysis of the synthetic substrate in a buffer pH 4.7, at the different GnCl molarities) was mirrored by monomer generation, seen by molecular sieve chromatography at the lower concentrations of the salt (GnCl 0.2 M for 30–35 min within the column). The dissociation was partly reversible and the yield of reversibility (standard conditions: 2 min in GnCl at the right conc., diluted to 0.001 M

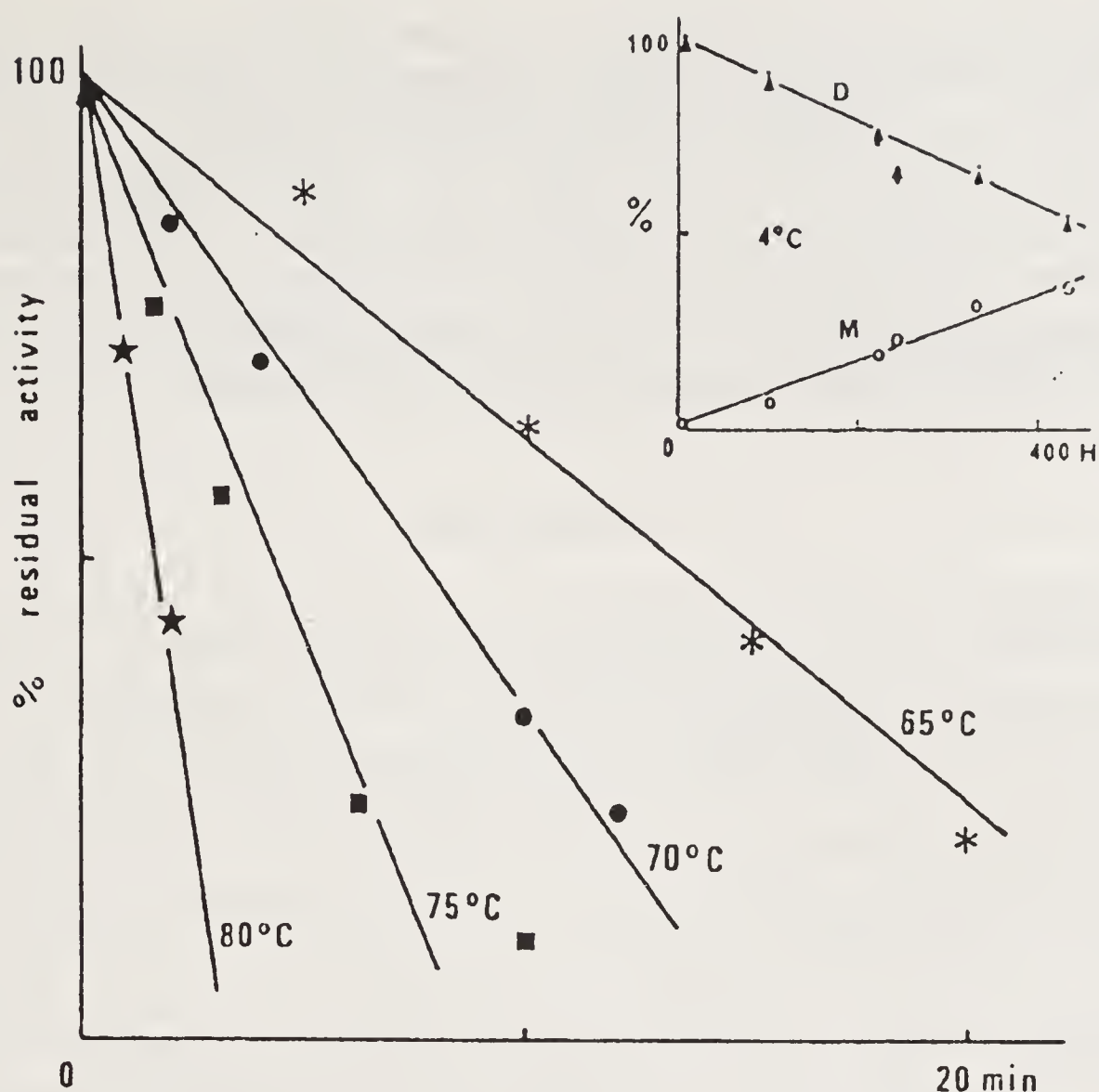


Figure 10. Temperature inactivation of the dimer. In the inset the dimer depolymerisation at 4°C, as a function of time.

Table 3. Effect of added chemicals on the dimer activity.

Chemical	Time (min)	Activity (%)	Reactivation
NaCl M		102	
Urea 0.1 to 3 M	5	98	
Triton X-100 1%	5	105	
10%	5	88	
	240	71	
20%	5	68	N.D.
Dioxane 15%	10	50	N.D.
GnCl 0.1 M	30	80	95
1.5 M	10	10	60
3 M	10	5	40
4.2 M	10	1-2	30

GnCl, pH 4.7 and analysed) was proportional to the initial GnCl concentration and to the time of contact. This system seems to obey a plain bimolecular mechanism, sometimes found in enzymes acting as dimer in "degenerated" conditions (43).

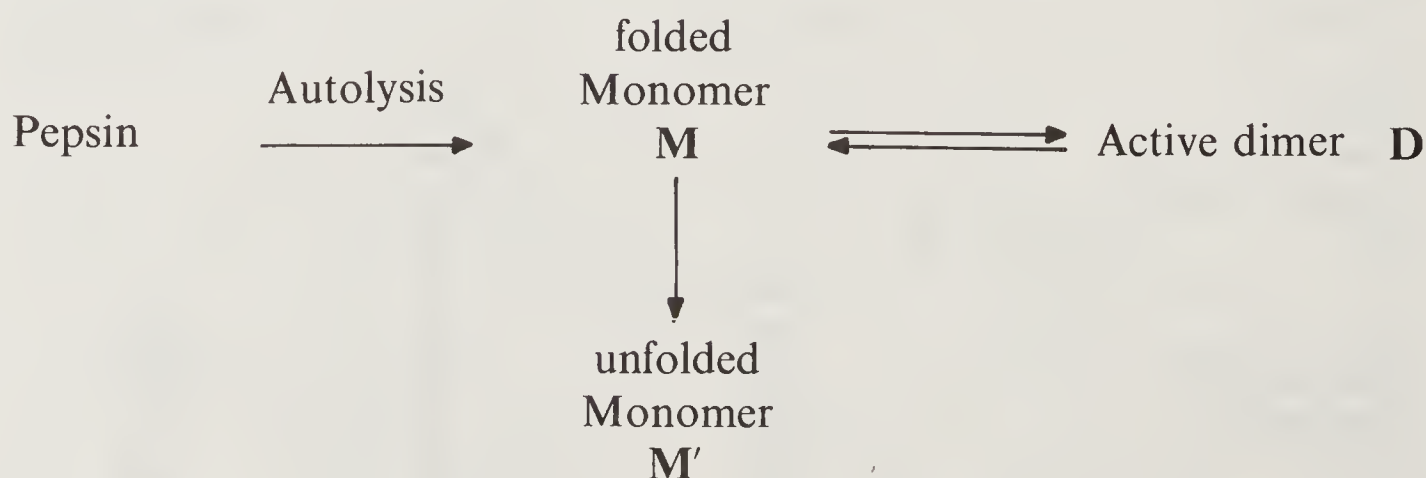
A general and probably oversimplified mechanism of constitution and reconstitution of the active species emerging from porcine pepsin autolysis could be:

The N-terminal domain of pepsin, cleaved by autolysis and able to escape through the dialysis bag, seems to be in a folded conformation (**M**) able to make up the active

dimer (**D**). Once formed the dimer is more resistant to the surrounding conditions but in equilibrium with the folded monomer: although more data is still needed, it seems however that once unfolded the monomer (**M'**) is unable to recover its native conformation and lead to the active dimer. This could be attributed, among other factors, to the presence of seven prolines in the sequence, which could hinder the capacity of the denaturated monomer (**M'**) to refold correctly due to the slow *cis-trans* isomerisations. Probably, this irreversible unfolding shifts the equilibrium towards an inactive and monomeric molecule under specific conditions.

Conclusion

These active fragments could be seen as models of retroviral aspartic proteases, which have been shown to act as dimers, and could also facilitate the search for new substrates and new inhibitors, provided the necessary adjustments for substrate specificity are subsequently made, a search sometimes hindered by the difficulty in obtaining sizeable quantities of pure material. Importantly, the evidence presented here shows that mono-domain aspartic proteases not only exist in retroviruses but that they can also be obtained by controlled cleavage of a bidomain aspartic protease. This poses the problem of their nomenclature. In fact there is no genetically coded "retrovirus protease" as such, there is only half of it and there are no "active" fragments from bilobal aspartic proteases. There are, in both cases, inactive molecules which gain their activity by dimerisation. These polypeptides, bearing only half of the two necessary active triads "Asp-Thr-Gly", have a catalytic power which is only potential until dimerisation occurs. Thus they are (and should be) called mono-aspartic-proteases with the indication of their origin in brackets. The fragment arising from porcine pepsin, presented in this work, is thus a "mono-aspartic-protease (porcine pepsin)".



References

1. Rich, D. H., Bernatowicz, M. S. & Schmidt, P. G. (1982) *J. Am. Chem. Soc.* **104**, 3535.
2. Blundell, T. & Pearl, L. (1989) *Nature (London)* **337**, 596–597.
3. Tang, J., James, M. N., Hsu, I. N., Jenkins, J. A. & Blundell, T. L. (1978) *Nature (London)* **271**, 618–621.
4. Pearl, L. H. & Taylor, W. (1987) *Nature (London)* **329**, 351–354.
5. Inouye, K. & Fruton, J. S. (1967) *Biochemistry* **6**, 1765–1777.
6. Nashed, N. T., Louis, J. M., Sayer, J. M., Wondrak, E. M., Mora, P. T., Oroszlan, S. & Jerina, D. M. (1989) *Biochem. Biophys. Res. Commun.* **163**, 1079–1085.

7. Moore, M. L., Bryan, W. M., Fakhouri, S. A., Magaard, V. W., Huffman, W. F., Dayton, B. D., Meek, T. D., Hyland, L., Dreyer, G. B., Metcalf, B. W., Strickler, J. E., Gorniak, J. G. & Debouck, C. (1989) *Biochem. Biophys. Res. Commun.* **159**, 420–425.
8. Mantle, D., Hardy, M. F., Lauffart, B., Mc Dermott, J. R., Smith, A. I. & Permington, J. T. (1983) *Biochem. J.* **211**, 567–573.
9. Pierzchala, P. A., Dorn, C. A. & Zimmerman, M. (1979) *Biochem. J.* **183**, 555–559.
10. Zimmerman, M., Ashe, B., Yurewicz, E. C. & Patel, G. (1977) *Anal. Biochem.* **78**, 47–51.
11. Stack, M. S. & Gray, R. D. (1989) *J. Biol. Chem.* **264**, 4277–4281.
12. Inouye, K., Voynick, I. M., Delpierre, G. R. & Fruton, J. S. (1966) *Biochemistry* **5**, 2473–2483.
13. Sanchez-Pescador, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempien, M. M., Brown-Shirmer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D. & Luciw, P. A. (1985) *Science* **227**, 484–492.
14. Jacks, T., Power, M. D., Marziak, F. R., Luciw, P. A., Barr, P. J. & Varmus, H. E. (1988) *Nature (London)* **331**, 280–283.
15. Moore, M. L., Bryan, W. M., Fakhouri, S. A., Magaard, V. W., Huffman, W. F., Dayton, B. D., Meek, T. D., Hyland, L., Dreyer, G. B., Metcalf, B. W., Strickler, J. E., Gorniak, J. G. & Debouck, C. (1989) *Biochem. Biophys. Res. Commun.* **159**, 420–425.
16. Nashed, N. T., Louis, J. M., Sayer, J. M., Wondrak, E. M., Mora, P. T., Oroszlan, S. & Jerina, D. M. (1989) *Biochem. Biophys. Res. Commun.* **163**, 1079–1085.
17. Bergman, D. & Kent, S., personal communication.
18. Navia, M. A., Fitzgerald, P. M. D., Mc Keever, B. M., Leu, C. T., Heimback, J. C., Heber, W. K., Sigal, J. S., Darke, P. L. & Springer, J. P. (1989) *Nature (London)* **337**, 615–620.
19. Weber, I. T., Miller, M., Jaskolski, M., Leis, J., Skalka, A. M. & Wlodawer, A. (1989) *Science* **243**, 928–931.
20. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. & Kent, S. B. H. (1989) *Science* **245**, 616–621.
21. Perlmann, G. E. (1954) *Nature (London)* **46**, 406.
22. Funatsu, M. & Tokuyatsu, K. (1959) *J. Biochem. (Tokyo)* **46**, 1441–1451.
23. Galen, F. X., Devaux, C., Guyenne, T., Ménard, J. & Corvol, P. (1979) *J. Biol. Chem.* **254**, 4848–4855.
24. Neurath, H. (1980) in *Protein Folding*, ed. Jaenicke, R. (Elsevier/North Holland, Biochem. Press, Amsterdam, New York) pp. 501–524.
25. Vita, C., Dalzoppo, D. & Fontana, A. (1985) *Biochemistry* **24**, 1798–1806.
26. Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zanai, M. & Zambonin, M. (1986) *Biochemistry* **25**, 1847–1851.
27. Neurath, H. (1987) in *Methods in Protein Sequence Analysis*, ed. Walsh, K. A. (Humana Press, Clifton N. J.), pp. 3–19.
28. Jacks, T., Power, M. D., Maziarz, F. R., Luciw, P. A., Barr, P. J. & Varmus, H. E. (1988) *Nature (London)* **331**, 280–283.
29. Coletti-Previero, M.-A. & Previero, A. (1989) *Anal. Biochem.* **180**, 1–10.
30. Bianchi, M., Boigegrain, R. A., Castro, B. & Coletti-Previero, M.-A. (1990) *Biochem. Biophys. Res. Commun.* **167**, 339–344.
31. Pugnière, M., Skalli, A., Coletti-Previero, M.-A. & Previero, A. (1986) *Proteins: Struc. Func. Genet.* **1**, 134–138.
32. Pugnière, M., San Juan, C., Coletti-Previero, M.-A. & Previero, A. (1988) *Biosci. Rep.* **8**, 263–269.
33. Favel, A., Mattras, H., Coletti-Previero, M.-A., Zwilling, R., Robinson, E. A. & Castro, B. (1989) *Int. J. Pept. Protein Res.* **33**, 202–208.
34. Hartsuck, J. A. and Remington, S. J. entry 1PSG of the Protein Data Bank from Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Schimanouchi, T. & Tasumi, M. J. (1977) *J. Mol. Biol.* **112**, 535–542.
35. Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D. & Davies, D. R. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7009.
36. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. & Kent, S. B. H. (1989) *Science* **245**, 616–621.
37. Miller, I. T., Jaskolski, M., Mohana Rao, J. K., Leis, J. & Wlodawer, A. (1989) *Nature (London)* **337**, 576–579.

38. Weber, I. T., Miller, M., Jaskolski, M., Leis, J., Skalka, A. M. & Wlodawer, A. (1989) *Science* **243**, 928–931.
39. Pflugrath, W., Sack, J. S., Saper, M. A. & Quiocho (1987) PS300 FRODO version 6.4.
40. Singh, U. C., Weiner, P. K., Caldwell, J. W. & Kollman, P. A. AMBER (UCSF), (1986) version 3.0. Depart. of Pharmaceutical Chemistry, University of California San Francisco.
41. Eisenberg, D. & McLachlan, A. D. (1986) *Nature (London)* **319**, 199–203.
42. Chiche, L., Gregoret, L. M., Cohen, F. E. & Kollman, P. A. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3240–3243.
43. Jaenicke, R. (1987) *Prog. Biophys. Mol. Biol.* **49**, 117–237.

Dynamic Aspects of Protein Specificity

G. C. K. Roberts

Department of Biochemistry and Biological NMR Centre, University of Leicester,
University Road, Leicester LE1 7RH, UK

Introduction

The last thirty years have seen an enormous increase in the structural information available about proteins, principally from X-ray crystallography but more recently also from nmr spectroscopy. This information has led to the formulation of 'rules' for protein structures, ranging from those determined by local stereochemical constraints, as embodied in the 'Ramachandran diagram', to the description of 'motifs' of super-secondary structures. However, like all molecules, proteins have structures which fluctuate at a wide variety of rates, and it is the combination of well-defined time-average structures with local flexibility which gives proteins some of their special, and biologically important, characteristics.

One of these characteristics is the ability of proteins to recognize other molecules with a high degree of specificity. Molecular recognition is, by its nature, a dynamic process; the equilibrium constant for the binding of a small molecule to a protein is not only a reflection of the change in free energy on binding, but is also the ratio of two rate constants. The biological functions of recognition require not only precise specificity but also a rapid response, and this combination of precision and speed seems to be achieved by making use of local flexibility.

In this paper, I shall describe some of our recent nmr work which has the aim of understanding the specificity of three proteins, dihydrofolate reductase, phospholipase A_2 , and the *trp* repressor, with the emphasis on the role of protein flexibility. This will also serve to illustrate the contribution nmr spectroscopy can make to the solution of these problems. Dihydrofolate reductase (dhfr), which catalyses the NADPH-dependent reduction of dihydrofolate and, to a lesser extent, folate, is the 'target' for the important 'anti-folate' drugs such as methotrexate, trimethoprim and pyrimethamine (1–4). Phospholipase A_2 (PLA₂; 5, 6), a calcium-dependent enzyme, catalyses the hydrolysis of the 2-acyl ester bond of phospholipids to release a fatty acid, notably arachidonic acid which is, in turn, the precursor of the prostaglandins, thromboxanes and leukotrienes, mediators of the inflammatory response. Tissue PLA₂ has therefore been identified as a potential target of anti-inflammatory agents. Finally, the *trp* repressor of *E. coli* (7–9) is a member of the "helix-turn-helix" family of DNA binding proteins; it shows specific, tryptophan-dependent, binding to the operator sequences of three operons, *trpEDCBA*, *aroH* and *trpR*, thus controlling the expression of the genes for the enzymes of tryptophan biosynthesis. These three small (M_r 14–25,000) proteins, for each of which high-resolution crystal structures are available, provide

convenient systems in which to attempt to obtain a quantitative understanding of the relation between structure, flexibility and specificity.

NMR Spectroscopy

The technical developments in nmr spectroscopy in recent years, notably in two-, three- and now four-dimensional nmr experiments and the use of stable isotope labelling, have dramatically increased its ability to provide structural and dynamic information on proteins in solution (10, 11). This has in turn increased its usefulness in the study of protein specificity. For example, with the development of site-directed mutagenesis to add to the skills of the organic chemist, it is now reasonably straightforward to make defined changes in the chemical structure of either the protein or the small molecule, and to measure the consequent changes in strength of binding. The functional (e.g., kinetic and binding) effects of amino-acid substitutions or of alterations in the chemical structure of the ligand can only be understood in the context of studies of their conformational effects and nmr has proved to be a convenient and sensitive way to study these (e.g., 12–19). Comparisons of the structures of mutant and wild-type enzymes and of complexes with different ligands, both by crystallography and by nmr have revealed considerable variation in the conformational consequences of chemical change, presumably arising from the differing degrees of flexibility of different parts of the structure. For example, crystallographic studies of the Asp 27 → Asn (D27N) mutant* of *E. coli* dhfr (20) and nmr studies of the corresponding *L. casei* mutant, D26N (15), show that any structural changes in the enzyme–methotrexate complex are very small and local (and may result from the displacement of a bound water molecule). By contrast, in the case of the T63Q substitution of dhfr, involving a residue which makes a hydrogen bond to the 2'-phosphate of the coenzyme, nmr analysis shows that the conformational effects include residues in the substrate binding site and are thus transmitted as much as 25 Å through the protein (19).

Nmr spectroscopy also provides a powerful method for the detection of individual interactions between protein and ligand in solution. Perhaps the most general approach to this is the use of intermolecular nuclear Overhauser effects (NOEs), which provide information on the spatial proximity of groups of the protein and of the ligand. For example, although a high resolution crystal structure of PLA₂ has been available for some time, it is only very recently that crystallographic information on inhibitor binding has been obtained (P. Sigler, personal communication). With the availability of a substantial number of resonance assignments in the proposed substrate binding site, we have begun to study the binding of a series of inhibitors of the enzymes (17 and W. U. Primrose, P. Slaich and G. C. K. Roberts, unpublished work): derivatives of *p*-bromophenacyl bromide, an irreversible inhibitor which reacts with His 48, and thioether or amide analogues of phosphatidylcholine. Clear intermolecular NOEs have been observed between the protons of bound inhibitors of all three classes

* The nomenclature used to describe mutants with single amino-acid substitutions is based on the single-letter code for the amino-acids; thus a mutant in which aspartic acid at position 27 of the wild-type sequence has been replaced by asparagine is denoted D27N. Unless otherwise specified, mutants discussed are those of *L. casei* dhfr.

and assigned protons of amino-acid residues – for example, between the aromatic protons of bound *p*-bromophenacyl bromide and protons of Phe 5, His 48, Tyr 69 and Phe 106. To allow us to assign the individual resonances of the hydrocarbon chains of the bound substrate analogues, we are synthesizing a series of specifically ^{13}C -labelled inhibitors. These are being employed in ^{13}C -edited NOESY experiments, in which one can selectively observe NOEs involving only the protons directly bonded to ^{13}C – hence picking out specifically the interactions of interest. These NOEs provide, for the first time, specific information on the location and orientation of these compounds in the binding site of the enzyme in solution. The results obtained with *p*-bromophenacyl bromide are at least qualitatively consistent with the recent crystal structure of *p*-bromophenacyl bromide-inhibited PLA_2 (21). It is clear that the use of ^{13}C -labelled ligands in conjunction with isotope-editing experiments will be a powerful approach to the identification of the intermolecular contacts in protein–ligand complexes, both in PLA_2 and more generally.

Similar kinds of isotope-editing experiments can be used to allow one to selectively observe the resonances of protons involved in hydrogen bonds between ligand and protein. For example, in dhfr the potent inhibitors methotrexate and trimethoprim bind with their 2,4-diaminopteridine or 2,4-diaminopyrimidine rings, respectively, protonated on N_1 (as shown by nmr; 22) and forming an ion-pair with the carboxylate of Asp 26. Specifically, the carboxylate of Asp 26 accepts hydrogen bonds from $\text{N}_1\text{-H}$ and the 2- NH_2 of the bound protonated inhibitor, while the second proton of the 2-amino group hydrogen bonds through a water molecule to Thr 116 (23). By using trimethoprim and methotrexate specifically ^{15}N -labelled at N_1 or the 2- NH_2 , together with isotope editing experiments, we have located the resonances of the protons involved in all three of these hydrogen bonds (24, and J. R. P. Arnold, J. DeGraw, J. Feeney and G. C. K. Roberts, unpublished work).

Nmr has proved particularly valuable in the study of structural *fluctuations*, as well as time-average structures. Motions in proteins and their complexes occur at rates spanning a wide timescale, from picoseconds to seconds, and nmr can provide detailed information over substantial parts of this range (25). Relaxation effects in ^1H , ^{13}C or ^{15}N spectra are sensitive to motions in roughly the $10^{11}\text{--}10^8\text{ s}^{-1}$ range. The microsecond range is a difficult one for nmr, but in the $10^2\text{--}10^4\text{ s}^{-1}$ range, analysis of effects on the resonance lineshape can provide quite precise estimates of rates. Finally, hydrogen exchange has provided valuable information on fluctuations occurring at $<10^2\text{ s}^{-1}$. Throughout these timescales, the major advantage of nmr is that it allows one to identify which residue(s) are involved in the rate processes being measured.

Structural Fluctuations in the Protein–Ligand Complex

The use of nmr in this area can be illustrated by our studies of the conformational fluctuations of the inhibitor trimethoprim (2,4-diamino, 5-[3',4',5'-trimethoxybenzyl]pyrimidine) bound to dhfr. Motions of the symmetrically substituted benzyl ring of trimethoprim were studied using [*m*-methoxy- ^{13}C]-trimethoprim (26). Relaxation measurements revealed fluctuations of $\pm 25\text{--}35^\circ$ at rates of about 10^9 s^{-1} , while lineshape analysis of the ^{13}C resonances shows that 180° 'flips' about the symmetry axis of the benzyl ring are much rarer, but still occur at 250 s^{-1} at room temperature.

A major contribution to the barrier to this 'flipping' comes from steric interactions *within* the trimethoprim molecule, between the benzyl and pyrimidine rings, and molecular mechanics calculations show that the 'flipping' of the benzyl ring must be accompanied by a change of 30–60° in the relative orientation of the two rings in bound trimethoprim. This in turn requires quite significant transient changes in the packing of amino-acid side-chains around the ring. The 'flipping' of symmetrically substituted aromatic rings in proteins, whether of ligand molecules (27) or of the tyrosine or phenylalanine residues of the protein itself, is a quite general observation (e.g., ref. 25). The rate is determined by the packing of amino-acid side-chains around the ring, and their ability to move (by about 1 Å) so as to allow the ring to flip – that is, by the local flexibility of the protein.

As noted above, in dhfr the bound protonated inhibitor forms hydrogen bonds and an ion pair with the carboxylate of Asp 26, and we have located the resonances of the protons involved in these hydrogen bonds. This allows us to measure the rate of exchange of these protons with the solvent and hence the rate of making and breaking of the hydrogen-bonds between inhibitor and enzyme. In the case of the trimethoprim N₁H–Asp 26 hydrogen bond, this rate is 34 s⁻¹ at room temperature (26), and clearly results from a fluctuation in the protein structure distinct from that involved in the ring flipping described above. It seems very likely that if both fluctuations happened to occur simultaneously, this would result in the dissociation of trimethoprim from the enzyme.

The rates of all these fluctuations in the structure of the enzyme–trimethoprim complex are affected by coenzyme binding (26), and also by amino-acid substitutions. Dhfr D26E is kinetically and structurally closely similar to the wild-type enzyme (13, 28), but this structural similarity is accompanied by clear differences in the dynamics of bound trimethoprim. The rates of both the 'flipping' of the benzyl ring and of the making and breaking of the hydrogen bonds are increased, by as much as 30-fold, in this mutant (13). It appears that structural fluctuations are significantly more sensitive to the effects of amino-acid substitutions than is the time-average structure.

Ligand-Induced Conformational Changes, Conformational Equilibria and Alternative Modes of Ligand Binding

It is by now a truism that the binding of a small molecule to a protein is very often (perhaps invariably) accompanied by a change in the conformation of the protein. In many instances, these conformational changes have now been characterised in some detail by crystallography or nmr. This kind of 'induced fit' (29) behaviour can of course be generally described in terms of an equilibrium between two (or more) conformations of the protein, one favoured in the absence of ligand and the other favoured in the complex. In a number of instances, when the interconversion rates are sufficiently slow, these conformational equilibria can be observed directly by nmr (e.g., 14, 15, 30–32). In addition, when the equilibrium constant is sufficiently close to one, it is possible for a given ligand to bind to the same site on a protein in more ways than one with very similar binding energy.

One kind of ligand-induced change which is commonly observed is a 'closing' of the protein around the ligand, which in some enzymes may have the function of

excluding water from the active site. Perhaps the best-known example of this kind of behaviour is the 'hinge' movement in hexokinase (33), but it is also seen, on a much smaller scale, in PLA₂. Bovine and porcine PLA₂ differ at 19 positions in the sequence, together with a single-residue insertion at position 120 of the porcine enzyme. The crystal structures of the two enzymes (34) indicate that these sequence differences result in generally small differences in three-dimensional structure, with the exception of the loop formed by residues 59–70. There is only a single residue difference in this loop (Val 63 to Phe) and yet conformations it adopts in the two crystal structures are such that the position of Tyr 69 differs by 5 Å. This difference has been implicated in the different micelle binding properties displayed by the bovine and porcine enzymes. However, such a structural difference between the two enzymes in solution is not apparent from our nmr studies. Indeed, analysis of the observed NOEs suggests that the *time-average* solution state conformations of the two enzymes are very similar in this region. The crystallographic and nmr observations can most easily be reconciled by suggesting that there is a degree of flexibility in this loop, and that it adopts different conformations in the two crystal structures not because of any intrinsic preference but due to the effects of crystal packing forces. Analysis of the changes in NOEs between residues in the active site on ligand binding shows that substrate analogue binding is accompanied by a movement of Tyr 69 so that its side-chain 'caps' the entrance to the active site (17); the conformational equilibrium of the flexible 59–70 loop is thus shifted towards one defined conformation on ligand binding.

The conformational equilibrium in the complex of *L. casei* dhfr with trimethoprim and NADP⁺ involves primarily the nicotinamide ring of the coenzyme (30, 31). In one conformation this ring is bound specifically to its site on the enzyme, but in the other, rotations about the nicotinamide ribose C5'–O and pyrophosphate O–P bonds have altered the conformation of the coenzyme so that the nicotinamide ring is no longer in contact with the enzyme, but is hanging free in solution, the coenzyme being bound solely through its adenosine moiety. These two rather different conformations are almost equally populated. The oxidised nicotinamide ring seems to contribute very little to the overall binding energy; it seems likely that the favourable interactions are balanced by the energetically unfavourable desolvation of the charged ring. Due to this fine balance, the relative populations of the two conformations are readily altered by changes in coenzyme or inhibitor structure or by amino-acid substitutions in the protein. For example, in the enzyme–trimethoprim–thioNADP⁺ complex or the dhfr W21L–trimethoprim–NADP⁺ complex *only* the conformation in which the nicotinamide ring is not in contact with the enzyme is detectable (13, 31).

The dhfr inhibitor methotrexate is a close structural analogue of the substrate folate, the important difference being a 4-amino rather than a 4-axo substituent on the pteridine ring of methotrexate. In spite of this close structural similarity, methotrexate is an inhibitor rather than a substrate, and binds some 10⁵ times more tightly than folate. Furthermore, the stereochemistry of reduction implies a difference of 180° in the orientation of the pteridine ring between substrate and inhibitor (35 and refs. therein). Nmr experiments have recently shown that both the enzyme–folate and the enzyme–folate–NADP⁺ complexes exist in solution as a mixture of slowly interconverting conformations whose proportions are pH-dependent (14 and references therein). The structure difference between these conformational states are localised to the active site region, close to the pteridine and nicotinamide rings, and NOE experiments have recently shown that one crucial difference between these conformations

is the orientation of the pteridine ring of the substrate. In the conformations denoted I and IIa, NOEs are observed between the pteridine 7-proton of bound folate and the two methyl groups of Leu 27, indicating that the ring has the same orientation in the binding site as that of methotrexate. In conformation IIb, by contrast, no such NOEs are observed; given the structural constraints of the pteridine ring binding site, this can only be explained if the ring has turned over by approximately 180° about an axis along the C2–NH₂ bond (14). Only the latter is a catalytically productive mode of binding for the substrate. In this case, therefore, the substrate is able to bind, with very similar affinity, in both productive and non-productive orientations. Studies of the D26N mutant (15) strongly suggest that Asp 26 is the group responsible for pH-dependence of this conformational equilibrium.

The *trp* repressor has been studied in detail by Sigler and his colleagues who have reported a number of high-resolution crystal structures (36–39). The protein is a dimer, each subunit consisting of six α -helices of which four interlock with the corresponding helices of the other subunit of the dimer to form the core of the protein and two form the “helix–turn–helix” DNA-binding domain. Tryptophan binds between the core and the helix–turn–helix domains. Both the crystallographic and nmr work suggests that the DNA-binding domains can adopt more than one orientation relative to the ‘core’ of the molecule (18, 38), and this may be influenced by tryptophan binding. It is particularly striking that, again both in the crystal and in solution, the corepressor tryptophan and the inducer indole-3-propionic acid bind to the protein at the same site but with their indole ring in a different orientation (18, 39). It remains to be established whether this difference is coupled to the equilibrium between the various orientations of the DNA-binding domain; this could in turn help to explain the markedly different effects of these two ligands on the affinity of the repressor for the operator.

Cooperativity in Ligand Binding

The phenomenon of ‘allostery’, in which the binding of a ligand to a protein at one site affects the binding of the same or a different ligand at a separate site, underlies many biologically important regulatory processes – for example, the control of gene expression in systems such as the *trp* repressor. To date, we have studied this in most detail in dhfr; in common with most dehydrogenases, the affinity of the enzyme for substrates and substrate analogues is markedly affected by the binding of coenzyme. In the case of dhfr, this cooperativity plays a role in catalytic turnover and in the mechanism of inhibition by antifolates. Thus, the inhibitor methotrexate binds to the enzyme some 600 times more tightly in the presence of the coenzyme than in its absence (40). Some contribution to this cooperativity no doubt comes from the van der Waals’ contact between the nicotinamide ring of the coenzyme and the pteridine ring of the inhibitor in the ternary complex. However, nmr has provided clear evidence that conformational effects are transmitted through the protein, and is allowing us to begin to define the route of transmission. The coenzyme-induced conformational changes appear to include an axial movement of helix C, residues 42–49, which runs from the binding site for the adenosine moiety of the coenzyme to that for the *p*-aminobenzoyl ring of the substrates and inhibitors (19, 41). The changes in packing around the benzoyl ring resulting from this conformational change are reflected in a 3-fold *increase* in the rate of ring flipping (27). Interestingly, in a mutant, dhfr T63Q, having a substitution in the binding site for the coenzyme 2'-phosphate there is clear

nmr evidence for structural changes in the inhibitor binding site, and indications from analysis of NOEs that these are transmitted through helix C (19).

In collaboration with Dr. C. Fierke and Prof. S. Benkovic (Pennsylvania State University) we have undertaken detailed studies of the kinetic mechanism of *L. casei* dhfr (28). The most striking feature of this mechanism is that, in the kinetically preferred pathway for product release, tetrahydrofolate (FH₄) dissociates only *after* NADPH has bound. This originates from the marked negative cooperativity in binding between FH₄ and NADPH; FH₄ dissociates 300 times faster from the E·NADPH·FH₄ complex than from E·FH₄. An interesting consequence of this mechanism is that the free enzyme is not on the kinetically preferred pathway. Very similar negative cooperativity between NADPH and the stable product analogue 5-formyl-FH₄ was earlier demonstrated by nmr and binding constant measurements (42). The natural 6*R*, α *S* diastereoisomer of 5-formyl-FH₄ has a binding constant of $1.3 \cdot 10^8 \text{ M}^{-1}$ for the enzyme alone, but it binds 3 times more weakly in the presence of NADP⁺ and 600 times more weakly in the presence of NADPH. It is notable that the single atom change between NADP⁺ and thioNADP⁺ converts negative into positive cooperativity. The negative cooperativity must result from ligand-induced conformational changes, and nmr studies show that these changes are qualitatively distinct from those associated with the positive cooperativity between NADPH and methotrexate (41, 42). Thus, notwithstanding the small size and apparent simplicity of dihydrofolate reductase, both large positive and large negative free-energy coupling between coenzyme and substrate analogue binding has been observed, each associated with a distinct conformational change in the protein.

Conclusions

It is apparent even from these very restricted examples that the adaptability and flexibility of protein conformation plays a crucial role in molecular recognition, and that to understand these processes we need a much more sophisticated model than a lock and key. It is also clear that nmr has much to contribute to the solution of these problems. It allows one to monitor individual protein–ligand interactions and to compare the structure of wild-type and mutant proteins in solution. Perhaps most importantly, it permits the *structural and kinetic* characterisation of dynamic processes in proteins and their complexes over a wide range of rates. The limitations of the technique, for example in the size of protein that can be tackled and in the quantitative structural interpretation of the data, are gradually becoming less restrictive, with improvements in relaxation and NOE analysis, the continuing development of higher-field spectrometers, and the use of ¹³C and ¹⁵N labelling. With the combination of X-ray crystallography, nmr spectroscopy and site-directed mutagenesis, we have an extraordinarily powerful set of tools for the analysis not only of protein structure but also of protein function.

Acknowledgements

Many people, whose names appear in the reference list, have made essential contributions to the experiments discussed briefly here; my thanks to all of them, and particularly to Jim Feeney, Berry Birdsall, Julie Andrews, John Arnold, Janette Thomas,

Eva Hyde, Vasudevan Ramesh, Bill Primrose, Julie Fisher and Lu-Yun Lian. Work at Leicester has been supported by SERC, MRC, the Wellcome Trust, the Leverhulme Trust, NEDO (Japan), Du Pont Central Research and Fisons Pharmaceuticals.

References

1. Roth, B. & Cheng, C. C. (1982) *Prog. Med. Chem.* **19**, 270–320.
2. Blakley, R. L. (1985) in *Folates and Pterins*, eds. Blakley, R. L. & Benkovic, S. J. (Wiley, New York), vol. 1, pp. 191–253.
3. Roberts, G. C. K. (1989) in *Protein Structure and Engineering*, ed. Jardetzky, O., Proc. NATO Advanced Study Institute (Plenum, New York), pp. 209–220.
4. Roberts, G. C. K. (1990) in *Chemistry and Biology of Pteridines 1989; Pteridines and Folic Acid Derivatives*, eds. Curtius, H. C. Ghisla, S. & Blau, N. (de Gruyter, Berlin), pp. 681–693.
5. Verheij, H. M., Slotboom, A. J. & de Hass, G. H. (1981) *Rev. Physiol. Biochem. Pharmacol.* **91**, 91–203.
6. Achari, A., Scott, D., Barlow, P., Vidal, J. C., Otwinowski, Z., Brunie, S. & Sigler, P. B. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52**, 441–452.
7. Rose, J. K. & Yanofsky, C. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3134–3138.
8. Bennett, G. N. & Yanofsky, C. (1978) *J. Mol. Biol.* **121**, 179–192.
9. Gunsalus, R. P. & Yanofsky, C. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7117–7121.
10. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acid*, (Wiley, New York), Wuthrich, K. (1990) in *Protein Structure and Engineering*, ed. Jardetzky, O., Proc. NATO Advanced Study Institute, (Plenum, New York), pp. 69–78.
11. Ernst, R. R., Bodenhausen, G. & Wokaun, A. (1987) *Principles of nuclear magnetic resonance in one and two dimensions* (Clarendon Press, Oxford).
12. Hammond, S. J., Birdsall, B., Feeney, J., Searle, M. S., Roberts, G. C. K. & Cheung, H. T. A. (1987) *Biochemistry* **26**, 8585–8590.
13. Birdsall, B., Andrews, J., Ostler, G., Tendler, S. J. B., Feeney, J., Roberts, G. C. K., Davies, R. W. & Cheung, H. T. A. (1989) *Biochemistry* **28**, 1353–1362.
14. Birdsall, B., Feeney, J., Tendler, S. J. B., Hammond, S. J. & Roberts, G. C. K. (1989) *Biochemistry* **28**, 2297–2305.
15. Jimenez, M. A., Arnold, J. R. P., Andrews, J., Thomas, J. A., Roberts, G. C. K., Birdsall, B. & Feeney, J. (1989) *Protein Engineering* **2**, 627–631.
16. Birdsall, B., Tendler, S. J. B., Arnold, J. R. P., Feeney, J., Griffin, R. J., Carr, M. D., Thomas, J. A., Roberts, G. C. K. & Stevens, M. F. G. (1990) *Biochemistry* in press.
17. Primrose, W. U., Magolda, R. L. & Roberts, G. C. K. (1990) in preparation.
18. Hyde, E. I., Ramesh, V., Frederick, R. & Roberts, G. C. K. (1990) in preparation.
19. Thomas, J. A. (1990) Ph.D. thesis, University of Leicester; Thomas, J. A., Arnold, J. R. P., Andrews, J., Feeney, J. & Roberts, G. C. K., in preparation.
20. Howell, E. E., Villafranca, J. E., Warren, M. S., Oatley, S. J. & Kraut, J. (1986) *Science* **231**, 1123–1128.
21. Renetseder, R., Dijkstra, B. W., Huizinga, K., Kalk, K. H. & Drenth, J. (1988) *J. Mol. Biol.* **200**, 181–188.
22. Roberts, G. C. K., Feeney, J., Burgen, A. S. V. & Daluge, S. (1981) *FEBS Lett.* **131**, 85.
23. Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C. & Kraut, J. (1982) *J. Biol. Chem.* **257**, 13650–13662.
24. Bevan, A. W., Roberts, G. C. K., Feeney, J. & Kuyper, L. F. (1985) *Eur. Biophys. J.* **11**, 211.
25. Jardetzky, O. and Roberts, G. C. K. (1981) *NMR in Molecular Biology* (Academic Press, New York).
26. Searle, M. S., Forster, M. J., Birdsall, B., Roberts, G. C. K., Feeney, J., Cheung, H. T. A., Kompis, I. & Geddes, A. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3787–3791.
27. Clore, G. M., Gronenborn, A., Birdsall, B., Feeney, J. & Roberts, G. C. K. (1984) *Biochem. J.* **217**, 659.

28. Andrews, J., Fierke, C. A., Birdsall, B., Ostler, G., Feeney, J., Roberts, G. C. K. & Benkovic, S. J. (1989) *Biochemistry* **28**, 5743–5750.
29. Koshland, D. E. jr. (1958) *Proc. Natl. Acad. Sci. U.S.A.* **44**, 98–104.
30. Birdsall, B., Bevan, A. W., Pascual, C., Roberts, G. C. K., Feeney, J., Gronenborn, A. & Clore, G. M. (1984) *Biochemistry* **23**, 4733–4742.
31. Gronenborn, A., Birdsall, B., Hyde, E. I., Roberts, G. C. K., Feeney, J. & Burgen, A. S. V. (1981) *Mol. Pharmacol.* **20**, 145–153.
32. Fisher, J., Primrose, W. U., Roberts, G. C. K., Dekker, N., Boelens, R., Kaptein, R. & Slotboom, A. J. (1989) *Biochemistry* **28**, 5939–5946.
33. Bennett, W. S. & Steitz, T. A. (1980) *J. Mol. Biol.* **140**, 211–230.
34. Dijkstra, B. W., Renetseder, R., Kalk, K. H., Hol, W. G. J. & Drenth, J. (1983) *J. Mol. Biol.* **168**, 163–179.
35. Charlton, P. A., Young, D. W., Birdsall, B., Feeney, J. & Roberts, G. C. K. (1985) *J. Chem. Soc. Perkin Trans. I* 1349–1353.
36. Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L. & Sigler, P. B. (1985) *Nature (London)* **317**, 782–786.
37. Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988) *Nature (London)* **335**, 321–329.
38. Lawson, C. L., Zhang, R.-G., Schevitz, R. W., Otwinowski, Z. & Sigler, P. B. (1988) *Proteins* **3**, 18–31.
39. Lawson, C. L. & Sigler, P. B. (1988) *Nature (London)* **333**, 869–871.
40. Birdsall, B., Burgen, A. S. V. & Roberts, G. C. K. (1980) *Biochemistry* **19**, 3723–3731.
41. Hammond, S. J., Birdsall, B., Searle, M. S., Roberts, G. C. K. & Feeney, J. (1986) *J. Mol. Biol.* **188**, 81–97.
42. Birdsall, B., Hyde, E. I., Burgen, A. S. V., Roberts, G. C. K. & Feeney, J. (1981) *Biochemistry* **20**, 7186.

Protein–Carbohydrate Interactions – A Theoretical Approach

V. S. R. Rao

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

Why Computer Simulation?

All biochemical reactions involve the interaction of two or more molecules: at least one of them will have a specific recognition or binding site to which the other binds for the biochemical reaction to take place. The control of these biochemical reactions depends on the high degree of specificity of these molecular binding sites. Information about the specificity of the binding sites is essential for the understanding of biomolecular interactions. In other words, a knowledge of the three-dimensional structure of biomolecules and their complexes is extremely important in understanding various biological processes. Among the various methods that are being used today, perhaps the X-ray diffraction technique is the only method that gives information about the precise position of atoms and hence about the three-dimensional structure of proteins and other molecules. Although an understanding of the size and structure of a binding site can be had only by X-ray crystallographic studies, one major disadvantage of this method is that it provides information about a system which is frozen in a particular conformation. Hence the conformational requirement of an inhibitor or substrate to initiate the binding process cannot be established directly by this method. Moreover, it is often difficult to obtain suitable samples for this technique besides the technique being time-consuming and expensive. Hence most of these studies are aimed at solving the structure of a protein in the native state or with one or two inhibitors at the most. Such information is generally insufficient to understand or explain any biological reaction satisfactorily. Moreover, it has also been recently noted that unless the structure is solved to a high resolution (below 2 Å) the information derived from this technique about protein–ligand interactions is not always reliable (1). Because of these limitations, the wealth of data provided by X-ray crystallographic studies on protein structures has not been fully exploited in understanding biomolecular interactions. Hence the major tasks in computer modelling are:

1. To improve the X-ray data obtained at about 3 Å.
2. To generate new complexes to explain the available experimental data and specificity
 - (a) when information about only the approximate location of the ligand is available and
 - (b) when information about the native structure alone is known.

With the advent of fast computers, molecular modelling studies (2–14) have, in fact, made significant advances towards our understanding of conformation and interaction of biomolecules and opened new vistas in molecular biology. Prof. G. N. Ramachandran is a pioneer in this area and has made outstanding contributions. It is worth mentioning that some of the basic rules which Prof. Ramachandran proposed (15) to work out the allowed conformations of biomolecules are still widely used in computer modelling studies. The author had the privilege of associating with him for over a decade and considers it a great honour to contribute to this special volume dedicated to him brought out by the Indian Academy of Sciences, Bangalore.

Numerous papers have also been published on the determination of the most favoured conformation of biomolecules (substrates or inhibitors) by experimental and energy minimisation studies with a view to relate their shapes to the biological properties. The conclusions drawn from such correlations are suspect since we know from computer simulations of protein–ligand interactions that the global minimum energy conformation need not necessarily initiate the binding process. The conformation of the molecule which initiates the binding process should have either full or partial complementarity to the active site of the protein. The former leads to the ‘Lock and Key’ model (16). When the complementarity is partial, either the inhibitor or the enzyme, or both undergo significant conformational changes for formation of the final complex. This suggests that in the induced fit type of binding there exist intermediate stages wherein conformation of the ligand initiating the binding process is different from that in the final complex. Because of this complementarity requirement even small changes in the ligand make them bind in a flipped orientation (6–8). Hence it is not an exaggeration to say that molecular fit is the key factor in molecular recognition. The present article deals with the recent developments in computer simulations of three-dimensional models of proteins and their complexes with carbohydrates.

L-Arabinose Binding Protein

Silhavy *et al.*, (17) isolated from gram-negative bacteria a few periplasmic proteins specific for carbohydrates. These proteins serve as essential components of the osmotic shock-sensitive active transport system. They apparently mediate a recognition process in the periplasmic space that facilitates specific nutrient passage to the cytoplasmic membrane and subsequently to the cell.

L-Arabinose binding protein (ABP) was isolated from *E. coli* by Hogg and Englesberg (18). It is composed of a single polypeptide chain and contains 306 amino acid residues. It binds to L-arabinose (L-ara), D-galactose (D-gal) and to a less extent to L-fucose (L-fuc) and D-xylose (D-xyl) (figure 1). However it does not bind to D-glucose (D-glc). Quioco and coworkers determined the three-dimensional structure of ABP–L-arabinose complex at 2.8, 2.4 and 1.7 Å resolution (1, 19, 20). It was only at 1.7 Å resolution that the authors were able to determine the binding modes of both the α - and the β -anomers of L-ara unambiguously. The proposed hydrogen bonding scheme between the protein and the ligand also differed significantly in each study (table 1). Quioco and coworkers (1) therefore suggested that unless the structure is refined at high resolution (below 2 Å), the interactions derived from X-ray techniques are not always reliable. This questions the usefulness of the X-ray data obtained at

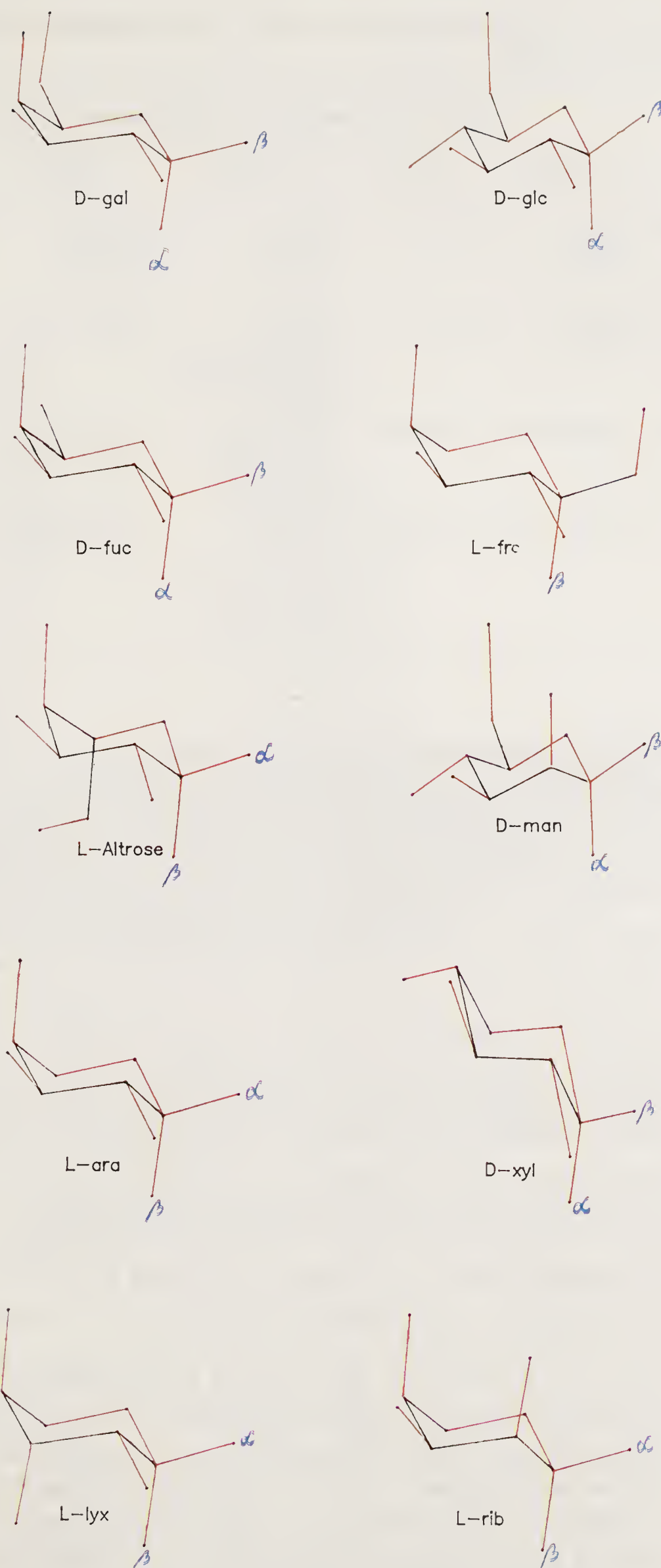


Figure 1. Perspective diagrams of the sugars (D-gal, D-fuc, D-glc, D-man, L-ara, D-xyl, L-lyx, L-rib, L-frcP and L-alt).

Table 1. Hydrogen bonding interactions in ABP-β-L-ara complex as proposed from various studies (1, 12, 19, 20).

Nature of study	Possible hydrogen bonds with the sugar hydroxyls				
	OH(1)	OH(2)	OH(3)	OH(4)	O5
X-ray studies at 2.8 Å resolution	Lys 10 Asp 89	Lys 10	Glu 14 Asn 205	Asn 205 Asn 232	—
X-ray studies at 2.4 Å resolution	Lys 10 Asp 90	Lys 10	Glu 14 (Possible through H ₂ O) Asn 205	Asn 232	—
X-ray studies at 1.7 Å resolution	Asp 90 Oδ2	Lys 10Nζ	Glu 14 Oε2 Asn 205 Nδ2 Asn 232 Nδ2	Asn 232 Oδ1 Arg 151 NH2	Arg 151 NH1
Theoretical studies	Asp 89 Oδ2	Lys 10Nζ	Glu 14 Oε2 (Possible through H ₂ O) Asn 205 Nδ2 Asn 232 Nδ2	Asn 232 Oδ1 Arg 151 NH2	Arg 151 NH1

resolutions of 2.4 Å and above. Mukhopadhyay and Rao (12) simulated the ABP-L-ara complex using the 2.4 Å resolution data (available at that time) in order to obtain better information about protein-carbohydrate interactions. Since the coordinates of the sugar were not deposited in the protein data bank, Mukhopadhyay and Rao (12) simulated the ABP-β-L-ara complex starting with the coordinates of the protein alone. Using contact criteria, these authors first constructed three-dimensional steric maps. These maps indicate that there are four regions that are allowed for β-L-ara to reach the binding site of the protein (figure 2). These allowed orientations were used as starting points in subsequent energy minimisation studies to determine the most probable orientation of the sugar in the binding site. The hydrogen bonding scheme in the most probable orientation is given in table 1 together with the data reported from X-ray crystallographic studies. It is gratifying to note that although the starting coordinates for these simulations were taken from the 2.4 Å resolution X-ray crystallographic data, the predicted hydrogen bonding scheme agrees well with the data reported at 1.7 Å resolution. Arg 151 and Asn 232 are predicted to form two hydrogen bonds each, the former with OH(4) and O5 and the latter with OH(3) and OH(4) which is in agreement with the high resolution data. These hydrogen bonds, excepting the one between OH(4) and Asn 232, were not considered in the interpretation of the 2.4 Å resolution data. During energy minimisation, several of the side chains of the protein moved significantly (table 2) from the positions assigned from the 2.4 Å resolution data. These changes lead to the possibility of extra hydrogen bonds which were in fact missing in the 2.4 Å resolution data. These results thus suggest that the positions of these side chains were not properly located from the 2.4 Å electron density map. The stereo projection of the most probable mode of binding of β-L-ara to ABP is displayed in figure 3.

Computer modelling studies also suggested a functional role for Asp 89. Though from the X-ray data in high resolution a hydrogen bond was predicted between OH(1)

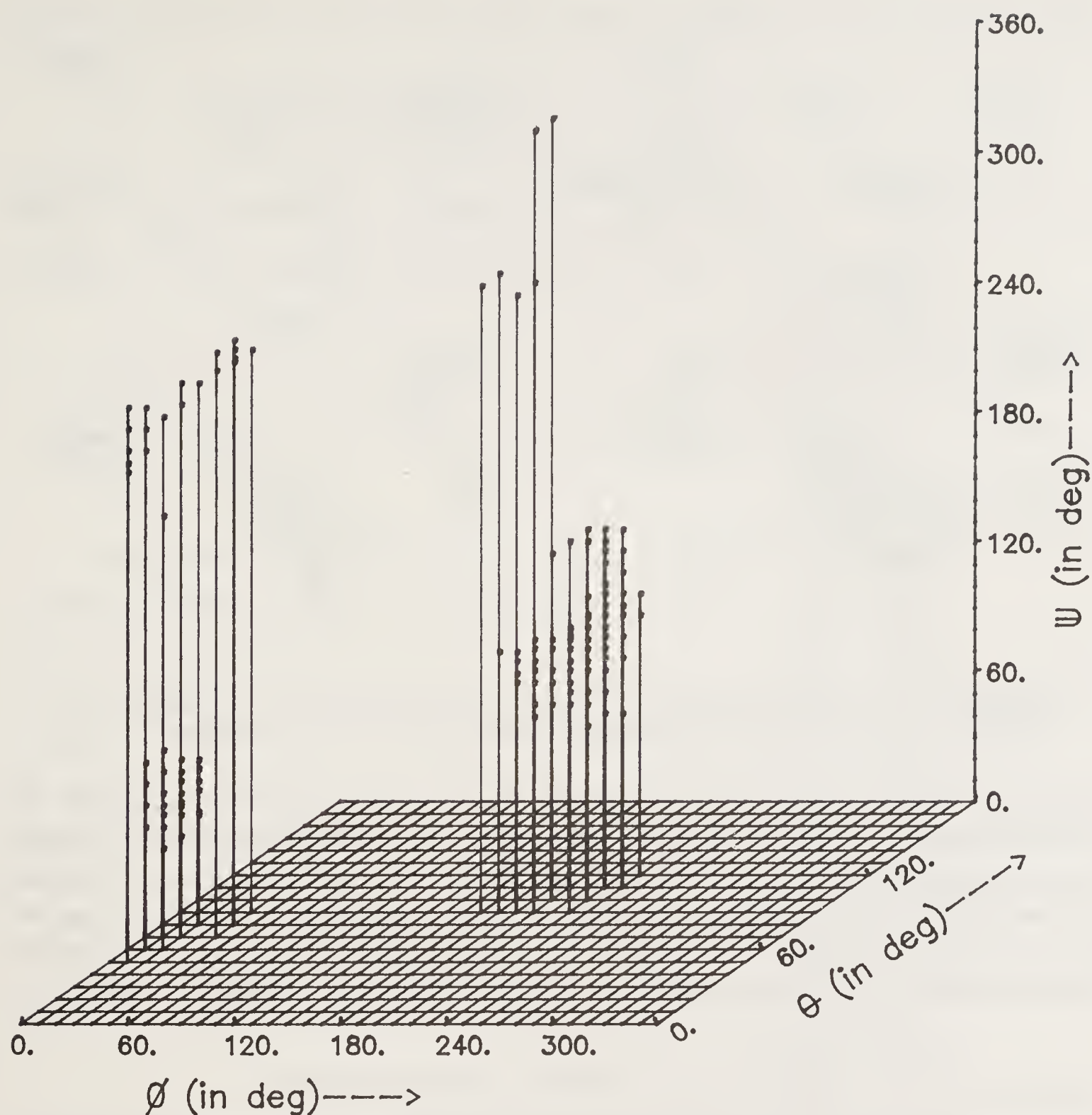


Figure 2. Steric map representing the allowed orientations for L-ara. The Eulerian angles ϕ , θ and ψ can be read in the following way: The length of the line joining the dot (·) and the point on the grid gives the value of the angle ψ . The grid point gives the values of ϕ and θ .

Table 2. Torsion angles of the side chains of the binding site residues obtained after minimisation (23).

Residues	(in degrees)			
	χ^1	χ^2	χ^3	χ^4
Lys 10	102	184	155	173
Arg 151	-33	-144	-54	-110
Met 204	185			
Asn 205	-105	21		
Asn 232	-110	32		

Torsion angles of the side chains of those residues not indicated did not change significantly from the 2.4 Å resolution data.

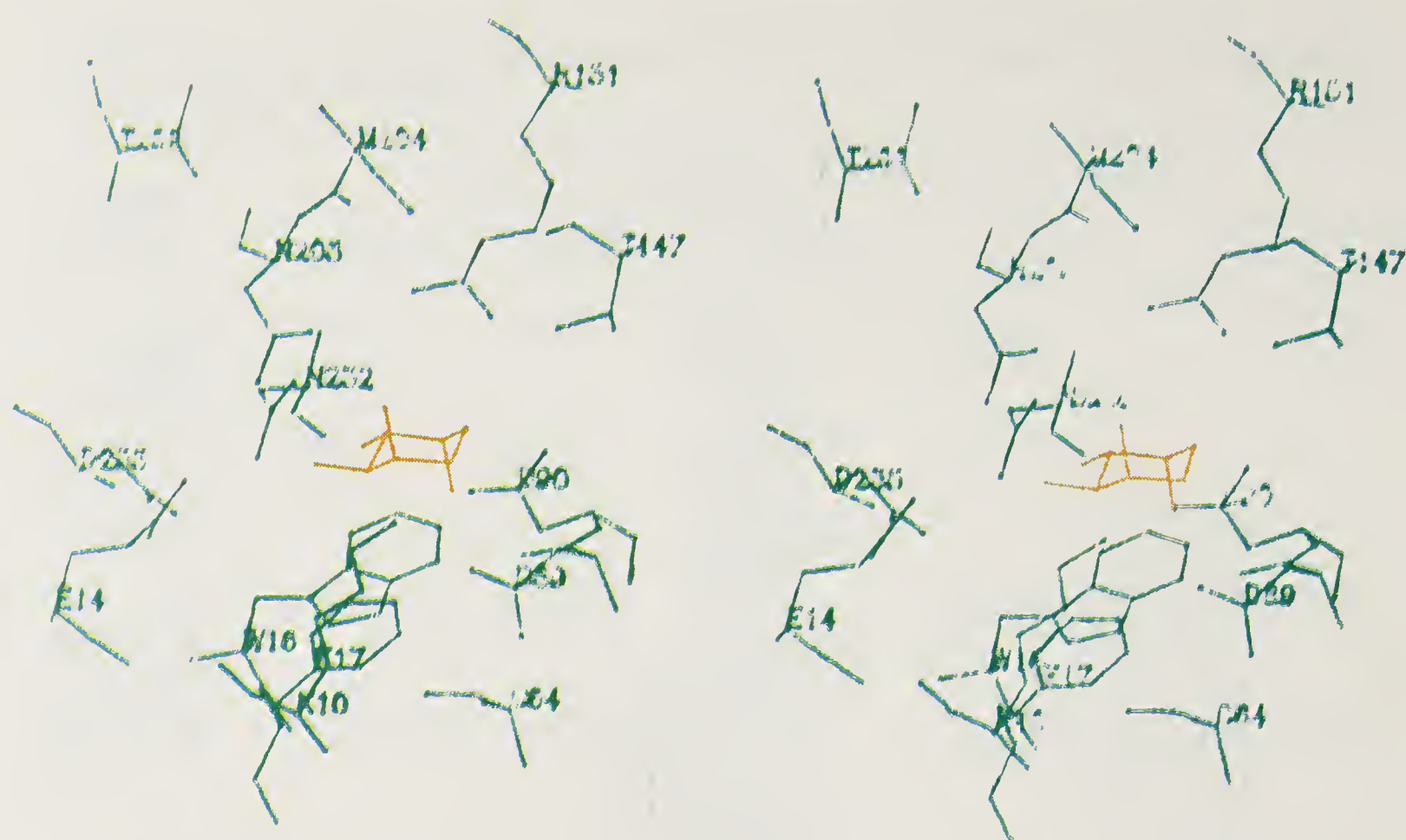


Figure 3. Stereoscopic projection of β -L-ara in the binding site of ABP in its minimum energy orientation. Hydrogen bonds (table 1) are not indicated.

and Asp 90, computer modelling studies favoured a hydrogen bond with Asp 89. This is interesting because this residue is conserved in two other sugar binding proteins, ribose and galactose (21). From the close agreement between the molecular modelling data and high resolution X-ray refinement data Mukhopadhyay and Rao suggested that the present computer modelling can be as effective in predicting the molecular details of protein–ligand interactions as the high resolution refinement studies.

Specificity of L-Arabinose Binding Protein

Mukhopadhyay and Rao (22, 23) have also studied the possible modes of binding of various sugars which differ in the nature and orientation of the substituents at different carbon atoms in the pyranose ring. These include the α - and β -anomers of D-gal, D-fuc, D-glc, L-ribose (L-rib), D-xyl, L-lyxose (L-lyx), L-altrose (L-alt) and L-fructose (L-frc). The most probable modes of binding of these sugars and the possible hydrogen bonding scheme arrived at theoretically are shown in table 3. The conformational energy of ABP– α -D-gal complex is the same as that of ABP– β -L-ara complex while that of ABP– α -D-fuc and ABP– α -D-xyl are much higher. Thus theoretical predictions are in agreement with the experimental results that D-gal binds to ABP with an affinity equal to that of L-ara and that D-fuc and D-xyl bind with a lower affinity (20, 24). A quick look at table 3 also reveals that the differences in the nature and orientations of the substituents at different ring carbons affect the orientation of the sugar in the binding site of the protein and these differences affect the hydrogen bonding scheme and hence the binding affinity of the sugars to ABP. The bidentate hydrogen bonds between Arg 151 and Asn 232 of the protein and the sugar are possible only in ABP–L-ara and ABP–D-gal complexes. In the other complexes of ABP with the moderate inhibitors (fucose and xylose) or the weak inhibitors (glucose) as ligands either or both of these bidentate hydrogen bonds are lost.

Table 3. Comparison of the most probable modes of binding of various sugars to ABP (22, 23).

Ligand	Orientation of Ligand						Confor- mational energy (kcal/mol)	Hydrogen bonding scheme
	ϕ (Degrees)	θ	ψ	X (Angstroms)	Y	Z		
β -L-ara	196	90	95	14.2	56.6	53.6	− 59.2	OH1–O δ 2(89) OH2–N ζ (10) OH3–N δ 2(205), N δ 2(232) OH4–NH2(151), O δ 1(232) O5–NH1(151)
α -D-gal	196	86	90	13.8	56.3	53.9	− 59.6	OH1–N ζ (10), O δ 2(89) OH2–N ζ (10) OH3–N δ 2(205), N δ 2(232) OH4–NH2(151), O δ 1(232) OH6–NH1(151)
α -D-fuc	192	88	99	13.8	56.3	53.9	− 57.7	OH1–O δ 2(89) OH2–N ζ (10) OH3–N δ 2(205) OH4–NH2(151), O δ 1(232)
α -D-glc	202	92	70	14.7	56.6	54.7	− 52.0	OH1–N ζ (10), O δ 2(89) OH2–N ζ (10) OH4–O δ 2(89)
α -D-xyl	200	49	140	14.0	56.9	53.5	− 54.2	OH1–O δ 2(89) OH2–N ζ (10) OH3–O δ 1(232) OH4–N δ 2(232) O5–NH1(151)
α -L-lyx	161	82	115	13.6	56.4	53.6	− 52.5	OH1–O(89) OH2–N ζ (10) OH3–O δ 1(205) OH4–NH2(151), O δ 1(232)
β -L-rib	163	100	120	14.1	56.7	53.8	− 49.8	OH1–O δ 2(89) OH2–NH1(151) OH3–O δ 1(205) OH4–NH2(151), O δ 1(232)
β -L-frcP	198	90	93	14.1	56.7	53.6	− 51.0	OH1–O δ 2(89) OH2–N ζ (10) OH4–NH2(151), O δ 1(232) O5–NH1(151)
α -L-alt	227	72	80	14.4	57.2	53.7	− 41.0	OH2–N ζ (10) OH4–O δ 1(205), NH2(151) O5–NH1(151)

For both L-ara and D-gal, though the anomer with the axial OH(1) is slightly more favoured than the other, both anomers form hydrogen bonds with identical amino acid residues in the binding site of ABP (12, 22). This suggests that the binding is not sensitive to the orientation of the hydroxyl group at the C-1 position. In other words ABP lacks anomeric specificity. On the other hand, introduction of a bulky substituent (e.g., a hydroxyl methyl group) at the C-1 position drastically reduces the binding affinity of the sugar, as has been found for L-frcP.

L-Rib and L-lyx differ from L-ara in the orientation of the hydroxyl groups at the C-2 and C-3 atoms respectively. Table 3 shows that a change in the orientation of the hydroxyl group at these positions from equatorial to axial severely affects the affinity of the sugars to the protein. D-Xyl, a pentose sugar is a moderate inhibitor to ABP, whereas D-glc, having an equatorial hydroxyl group at C-4 and a CH₂OH group at the C-5 position can act only as weak inhibitor which suggests that the unfavourable effect of an equatorially oriented –OH group is maximum when a –CH₂OH group is present on the adjacent C-5 atom. On the other hand, a sugar with an axial –CH₂OH group at C-5 (α -L-alt) may not bind to ABP at all.

Thus these computer modelling studies of ABP with different sugars not only explain the experimental observations but also give valuable information about the specificity of the protein.

Concanavalin A (Con A)

Con A binds to carbohydrate structures containing α -D-man and α -D-glc residues (25). The saccharide binding specificities of Con A are utilised for the isolation, purification, characterisation and sequencing of polysaccharides and glycoproteins. It is also widely used to probe changes in the structure, organization and dynamics of cell surface molecules during the development of non-oncogenic virus infections, oncogenic virus transformation, cell growth and cell division, cell movement and cell–cell interactions.

Con A exists as a dimer between pH 2.0 and 5.5 and as a tetramer at pH values above 5.5. The molecular weight of the monomer is about 26,000 daltons. Each Con A molecule has 237 amino acid residues and two divalent metal ions, Ca²⁺ and Mn²⁺. The three-dimensional structure of native Con A at 2.4 Å was reported by Hardman and Ainsworth as early as 1972 (26). These authors later crystallised a Con A– α -MeMan complex and solved the structure only at low resolution (6.0 Å) (27). These studies revealed that the carbohydrate binding site lies near Tyr 12, Asp 16, Tyr 100 and Asp 208 in a depression on the molecular surface of Con A, 10–12 Å away from the bound Mn²⁺ ion. But details of the carbohydrate–Con A interactions were not known. Using the information obtained on the native Con A at 2.4 Å resolution X-ray crystallographic data, Rao and coworkers have modelled the binding of a series of carbohydrate inhibitors to Con A [α - and β -anomers of D-glc (8, 28); α and β D-man (10, 14); α - and β -linked disaccharides (29, 30) and complex type glycopeptides (31, 32)] initially using contact criteria and later by energy minimisation. From steric maps (figures 4 and 5) these authors suggested that methyl β -D-glc can bind to Con A in three different modes (28) whereas methyl α -D-glc can bind in only one mode (8). In their most favourable modes of binding their alignment in the binding site of Con A differ (figures 6, 7). Methyl β -D-glc binds in a flipped or an inverted orientation compared to methyl α -D-glc (figure 7). These predictions are in agreement with the nmr studies (33). Computer modelling experiments also predicted that α -MeManP can reach the binding site of Con A in three different regions (figure 8). Of these, region I is favoured due to the possibility of forming better hydrogen bonds and hydrophobic interactions with the protein. Energy minimisation studies have revealed that two orientations around $\phi = 300 \pm 40$, $\theta = 130 \pm 40$ and $\psi = 160 \pm 20$ and $\phi = 330 \pm 10$, $\theta = 100 \pm 10$ and $\psi = 335 \pm 10$ are energetically favoured. The

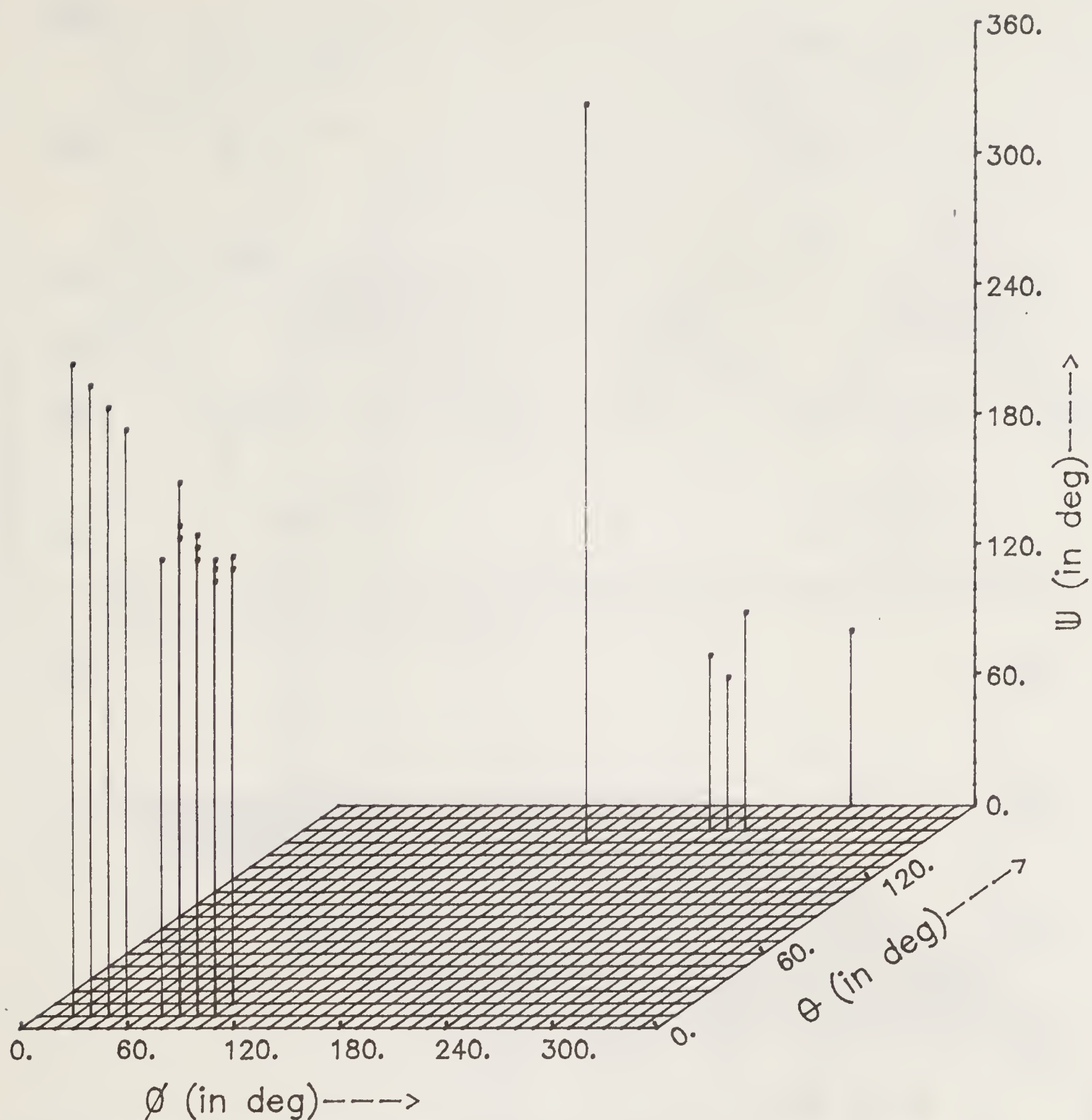


Figure 4. Steric map representing the allowed orientations of Me- β -D-glc in Con A binding site.

possible hydrogen bonds and the energy of these conformations calculated by Reddy and Rao are displayed in table 4 (34). It is interesting to note that the recently reported crystal structure of Con A- α -MeMan (35) agrees with the second conformation which has slightly higher energy. The energy difference between conformations I and II mainly arises due to the absence of hydrogen bonds involving C-2 hydroxyl group and O-5 atom of the sugar (table 4). In the crystal structure of the Con A- α -MeMan complex, the C-2 hydroxyl group forms an inter subunit hydrogen bond through a water bridge with the subunit of an adjacent tetramer. Since the energy calculations consider only a single monomer, the latter hydrogen bond is not accounted for. The possibility of the formation of such a hydrogen bond with the C-2 hydroxyl group may lower the total energy of the second conformation. This possibly accounts for the existence of this conformation in the solid state. It is also gratifying to note that the recently reported crystal structure of Con A- α -MeMan falls in the allowed region predicted (figure 8) using contact criteria (8).

Energy calculations of Reddy and Rao also include the binding modes of complex

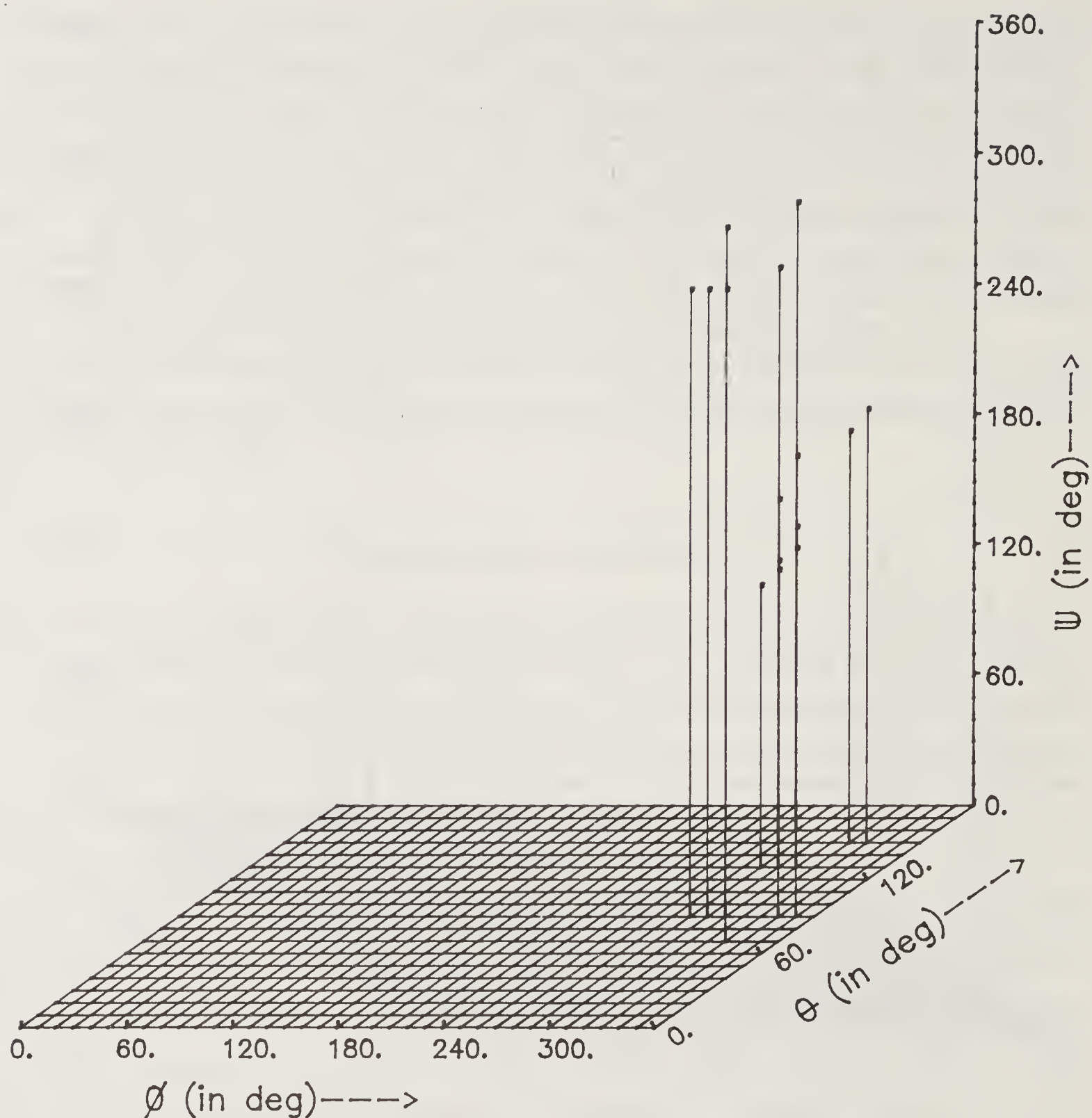


Figure 5. Steric map representing the allowed orientations of Me- α -D-glc in Con A binding site.

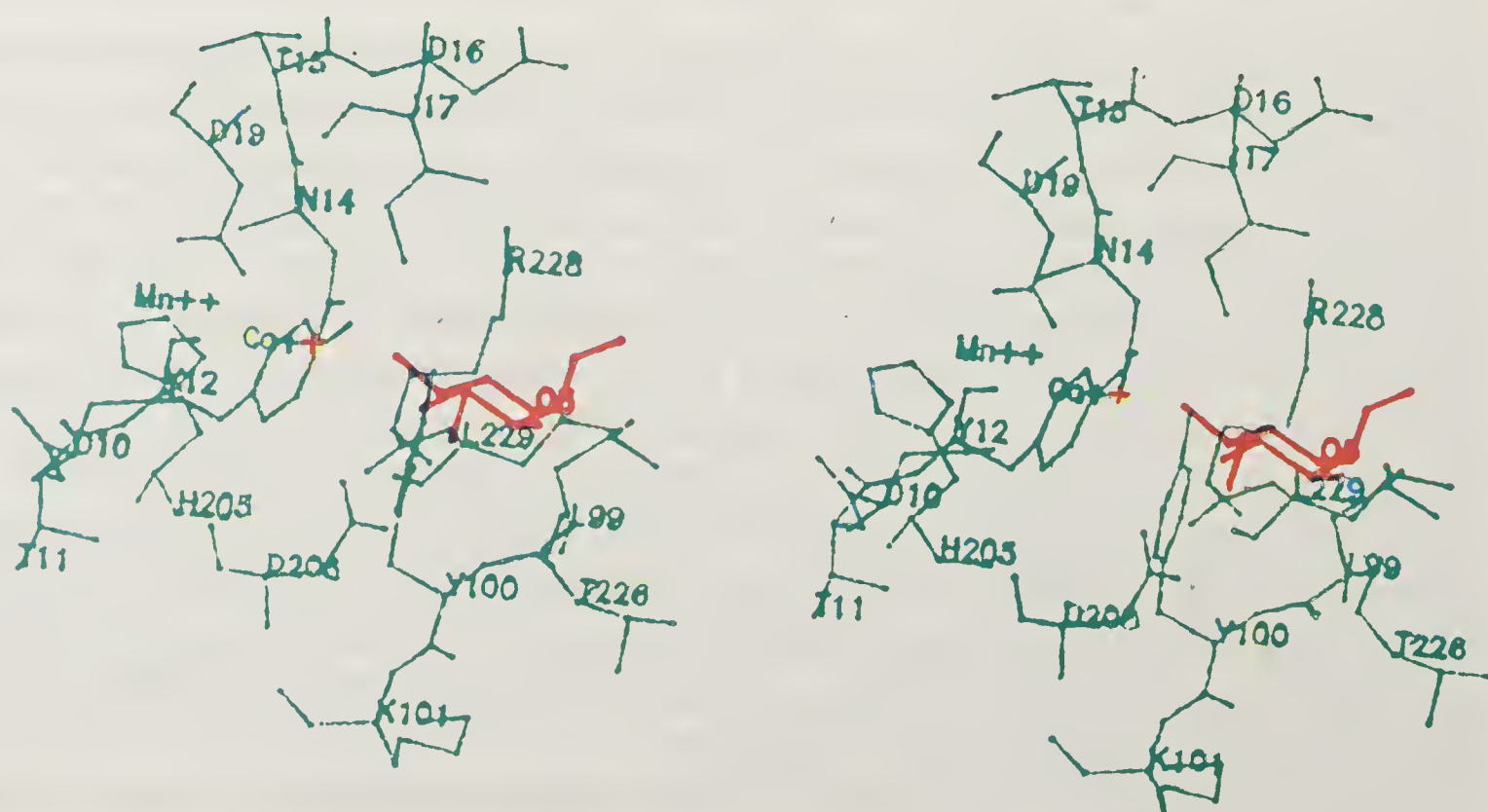


Figure 6. Stereoscopic projection of Me- α -D-glc in the binding site of Con A. Hydrogen bonds are not indicated.

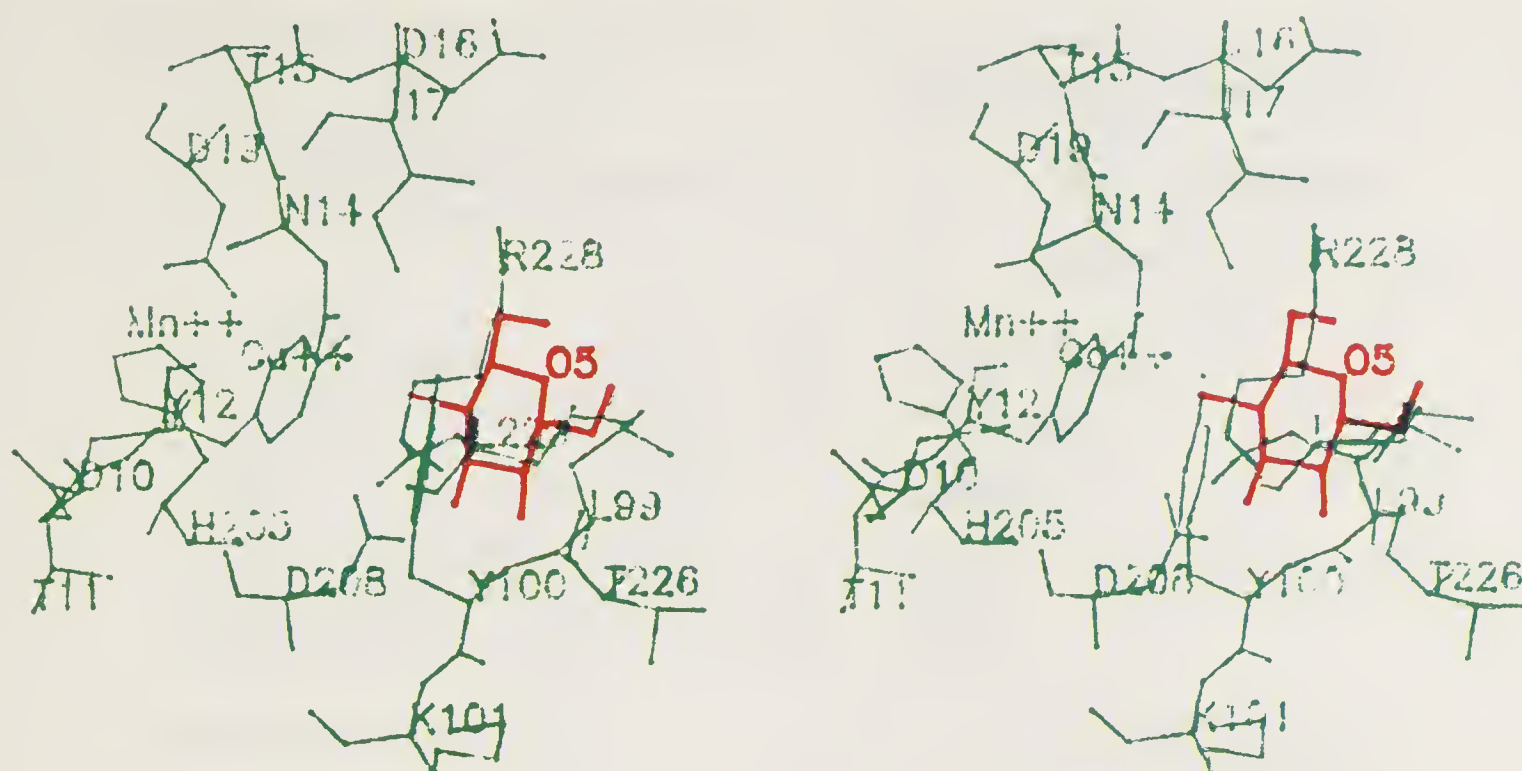


Figure 7. Stereoscopic projection of Me- β -D-glc in the binding site of Con A. Hydrogen bonds are not indicated.

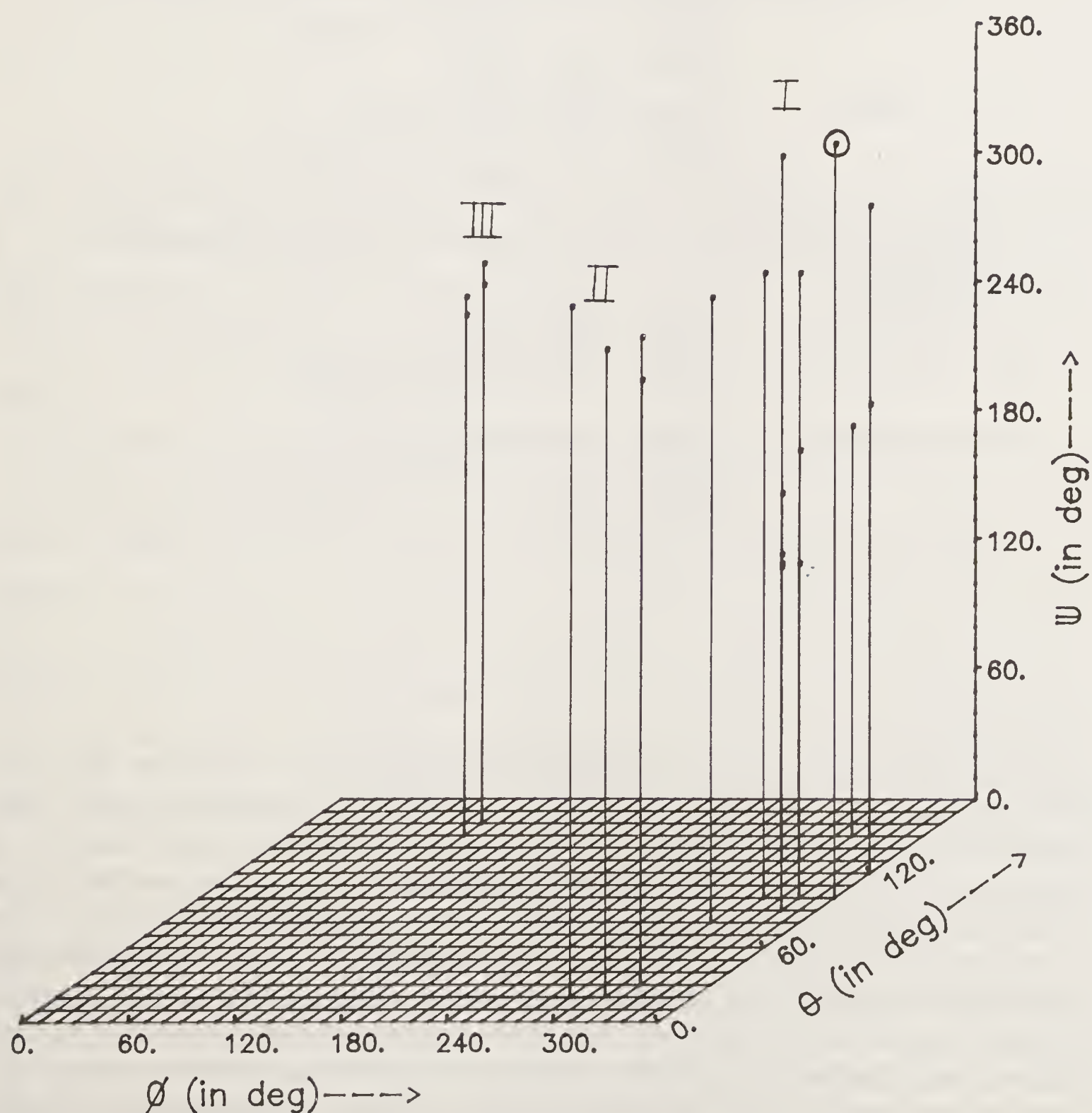


Figure 8. Steric map representing the allowed orientations of Me- α -D-man in Con A binding site. The orientation observed in the recent crystal structure study is marked (\odot).

Table 4. Conformational energies and hydrogen bonding scheme in the Con A–ligand complexes (34).

Normalised energies		Hydrogen bonding scheme	
1) Con A–a MeMan			
I) – 1076.6	OH2	Nδ2	Asn 14
	OH3	Nδ2	Asn 14, Oδ1 Asp 208
	OH4	NH	Leu 99, Oδ1 Asp 208
	O5	Nε	Arg 228
	OH6	O	Thr 226, Nε Arg 228
II) – 1072.2	OH3	NH	Arg 228, Nε Arg 228
	OH4	Oδ1	& Oδ2 Asp 208, NH Arg 228
	OH6	NH	Leu 99, NH Tyr 100, Oδ1 Asp 208
2) Con A–Trimannoside, (1–6) linked mannose (A) residues in the primary binding site.			
I) – 1085.2	OH3A	Oδ1	Asp 208
	OH4A	NH	Leu 99, Oδ1 Asp 208
	O5A	Nε	Arg 228
	OH6A	O	Thr 226, Nε Arg 228
	OH3C	OH	Tyr 100
II) – 1092.0	OH4A	Oδ1	& Oδ2 Asp 208
	OH6A	NH	Tyr 100, Oδ1 Asp 208
	OH2B	O	Pro 13
3) Con A–Trimannoside, (1–3) linked mannose (C) residue in the primary binding site.			
I) – 1078.5	OH2C	Nδ2	Asn 14
	OH3C	Nδ2	Asn 14
	OH4C	NH	Leu 99, Oδ1 Asp 208
	OH6C	O	Thr 226, Nε Arg 228
II) – 1084.0	OH4C	Oδ1	&Oδ2 Asp 208
	OH6C	NH	Tyr 100, Oδ1 Asp 208
	OH3A	Oδ1	Asp 16

carbohydrates to Con A (14). In simple trimannosidic core structures, the (1 \rightarrow 6) linked end mannose fits better than the (1 \rightarrow 3) linked end mannose at the sugar binding site of Con A (table 4). It is interesting to note (figure 9) that when (1 \rightarrow 6) end mannose is in the sugar binding site, the other two residues lie over the surface of the protein. In this the (1 \rightarrow 6) end mannose forms hydrogen bonds similar to that in conformation II of the α -MeMan. The C-2 hydroxyl of the middle residue forms a hydrogen bond with the backbone carbonyl group of Pro 13. There are no hydrogen bonds between the protein and the (1 \rightarrow 3) linked end mannose in this mode of binding. Preferred conformations of complexes of bigger carbohydrate structures with Con A have also been studied both by contact criteria and energy minimisation but these results are not discussed here as these are not solved by X-ray diffraction.

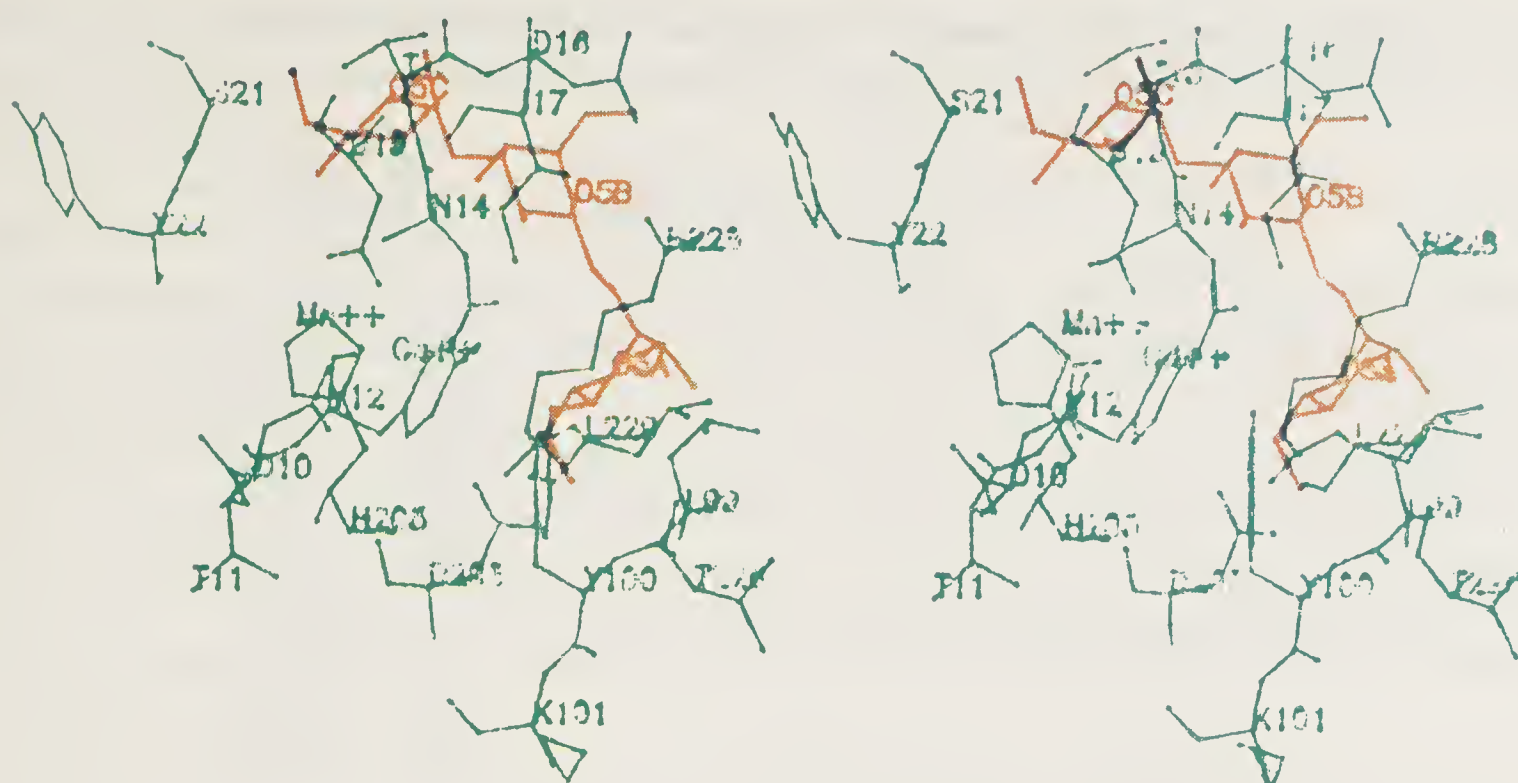


Figure 9. Stereoscopic projection of trimannosidic core structure with (1-6) man in the binding site of Con A. See table 4 for hydrogen bonding scheme.

Pea Lectin

Pea lectin (25) has a molecular weight of about 50,000 daltons. It also contains Mn^{2+} and Ca^{2+} ions and its carbohydrate binding specificity is very similar to Con A. Like Con A it favours to bind to α -anomers of D-man and D-glc and their α -linked oligosaccharides. A methyl substitution at C-3 or C-4 hydroxyl increases the binding of these monosaccharides to pea lectin. In this respect, pea lectin is dissimilar to Con A where the substitution of a methyl group at C-3 of D-glc drastically reduces its affinity.

The favoured modes of binding of methyl- α -D-glc and methyl 2,3 dimethyl- α -D-glc to pea lectin have been studied by Rao and coworkers (13). Methyl- α -D-glc can bind to pea lectin in three orientations of which $\phi = 250^\circ$, $\theta = 110^\circ$, $\psi = 120^\circ$ is most favoured (table 5). The first and third orientations favoured by Me α -D-glc in the binding site of pea lectin are very similar to the favoured orientations of Me α -D-glc in the Con A binding site. In these two orientations the sugar hydroxyls O3, O4 and O6 are involved in the formation of strong hydrogen bonds with the protein. In the second most favoured orientation the O6 hydroxyl group does not form any hydrogen bond with the protein. Computer modelling studies have also revealed that methyl groups can be added to O2 and O3 only in the third favoured orientation. In other words methyl 2,3-dimethyl- α -D-glc can bind to pea lectin in an orientation ($\phi = 355^\circ$, $\theta = 115^\circ$, $\psi = 330^\circ$) similar to the third in table 5. The first two orientations which are favoured for methyl- α -D-glc are completely eliminated for 2,3 substituted derivatives. In the orientation ($\phi = 355^\circ$, $\theta = 115^\circ$, $\psi = 330^\circ$) besides the hydroxyl groups at O3, O4 and O6 being involved in the formation of hydrogen bonds, the methyl group at C3 is placed close to a hydrophobic region constituted by Gly 97, Gly 98, Gly 99 and Tyr 100 and the hydroxy methyl group to Phe 123 leading to additional stabilising hydrophobic interactions. These results thus explain the hapten inhibition measurements on the precipitation reaction, between pea lectin and a phosphomannan which show that methyl 2,3-dimethyl- α -D-glc is a very potent inhibitor. It is interesting to

Table 5. Energy minimisation of the pea lectin–Me- α -D-Glc complex.

	Orientation of sugar						Energy (kcal/ mol)	Hydrogen bonding scheme
	ϕ (Degrees)	θ (Degrees)	ψ (Degrees)	X (Angstroms)	Y (Angstroms)	Z (Angstroms)		
1.	250	110	120	24.4	23.5	25.9	– 72.8	O3H...O ϵ 2(218) O4H...O δ 1(81) O6H...O δ 2(81)
2.	20	75	225	25.0	23.5	25.4	– 69.5	O2H...O δ 2(81) O3H...O ϵ 2(218) O4H...O ϵ 2(218)
3.	355	115	330	25.2	23.4	24.0	– 62.9	O3...NH(99) O4...NH2(125) O4H...O δ 1(81) O6H...O δ 2(81)

note that a similar methyl substitution on the O3 position of the sugar ligand drastically reduces its ability to bind to Con A.

Lysozymes

Lysozymes which are ubiquitous enzymes belong to the hydrolytic class of enzymes. They are defined as 1,4-beta-N-acetylmuramidases. They cleave the glycosidic bond between the C-1 atom of N-acetylmuramic acid (MurNAc) and the C-4 atom of N-acetylglucosamine (GlcNAc) in the bacterial cell wall peptidoglycan (36).

The pioneering crystallographic studies of Phillips and coworkers on hen egg white lysozyme have made it a model system for protein–carbohydrate interactions in general, and for enzyme–substrate interactions in particular (37). It is the first enzyme for which a detailed mechanism is proposed based on the crystal structure of an enzyme–saccharide inhibitor complex.

The main features of the mechanism are:

- (i) The enzyme can accommodate six sugar residues and the sugar binding sites are labelled as A, B, C, D, E and F.
- (ii) The glycosidic bond between saccharides occupying D and E sites is hydrolysed.
- (iii) Glu 35 acts as the proton donor. Asp 52 stabilises the oxocarbenium ion intermediate formed due to the protonation of the glycosidic oxygen and cleavage of bond between C-1 atom and the glycosidic oxygen.
- (iv) A sugar residue cannot occupy the D site in the normal chair conformation. It has to be distorted into a half-chair or sofa conformation. This geometric distortion or strain makes the conformation around C-1, C-2 and O-5 atoms planar and such planar conformation facilitates the formation of carbonium ion at C-1 atom in the transition state.

Most of the features of the above mechanism are supported by experimental evidence and the mechanism is generally held to be correct except for the role of the geometric

strain in the ground state. The main argument against steric strain in the ground state is that the enzyme structure is not rigid enough to cause a geometric or mechanical distortion of the substrate.

Many computer simulation studies have been carried out to clarify the role of the steric strain and the mechanism of catalysis.

Warshel and Levitt (38) used a hybrid method which couples a quantum mechanical method with an empirical force field to study the complex of (GlcNAc)₆ with lysozyme. They concluded that a sugar residue can be accommodated in D site without distortion and the driving force for catalysis is the electrostatic strain but not the geometric strain. The electrostatic strain is due to the binding of the saccharide residue in D site, which causes the desolvation of negatively charged carboxylic acid group of Asp 52. The electrostatic strain favours the formation of oxo-carbonium ion intermediate, since such an intermediate can form an ion-pair with the negatively charged carboxyl group of Asp 52.

Pincus and Scheraga (39) carried out potential energy calculations on lysozyme–saccharide complexes and showed that both GlcNAc and MurNAc can be accommodated in D site without distortion. They observed two modes of binding for (GlcNAc)₆. Depending on the orientation of the D and E rings in the active site, the two modes are termed as right and left side modes of binding. In the right side binding mode the saccharide ring in D site is located deep in the active site cleft and is near to the catalytic amino acid residues Glu 35 and Asp 52; whereas in the left side binding mode the saccharide residue in D site is somewhat removed from these two catalytic amino acid residues. These observations prompted them to suggest that right side binding mode is the catalytic mode of binding whereas the left side binding mode may be a productive mode but not a direct catalytic mode.

Post and Karplus (40) from a 55-ps molecular dynamics simulation of lysozyme–(GlcNAc)₆ complex also concluded that there is no ring distortion in the D site. During the molecular dynamics simulation, the chair form of the pyranose ring in site D remained unperturbed and the glycosidic dihedral angle O5–C1–O–C4' between rings D and E which was -54° in the initial structure, stabilised at -62° . The motions of the carboxyl group of Glu 35 led to hydrogen bonds with the endocyclic D ring oxygen O5 and the hydroxymethyl oxygen O6 but not with the glycosidic oxygen O. The closest approach of Glu 35 to O was greater than 4 Å with an improper orientation for hydrogen bond formation. Based on these results, Post and Karplus have suggested a new hydrolytic mechanism which does not require substrate distortion. In the first step of their mechanism the ring oxygen O5 is protonated causing cleavage of the C1–O5 bond and the formation of an acyclic oxo-carbonium ion intermediate. This intermediate is stabilised by Asp 52. Attack by water, cleavage of C1–O bond and ring closure leads to the products (figure 10).

In this respect it is interesting to note that the recent computer simulation studies of Lakshminarasimhulu and Rao (unpublished data) on T4 lysozyme–saccharide complexes have shown that all the saccharide ligands bind to T4 lysozyme in a mode analogous to the right side binding mode of hen egg white lysozyme and the mechanism of Post and Karplus may not be valid for this enzyme.

In conclusion, the molecular simulations allow us to study specificity, interactions of protein–carbohydrate complexes and also the mechanism of action of enzymes using the data obtained even at low resolution by X-ray diffraction studies on the complexes as well as native proteins. Thus computer modelling studies complement

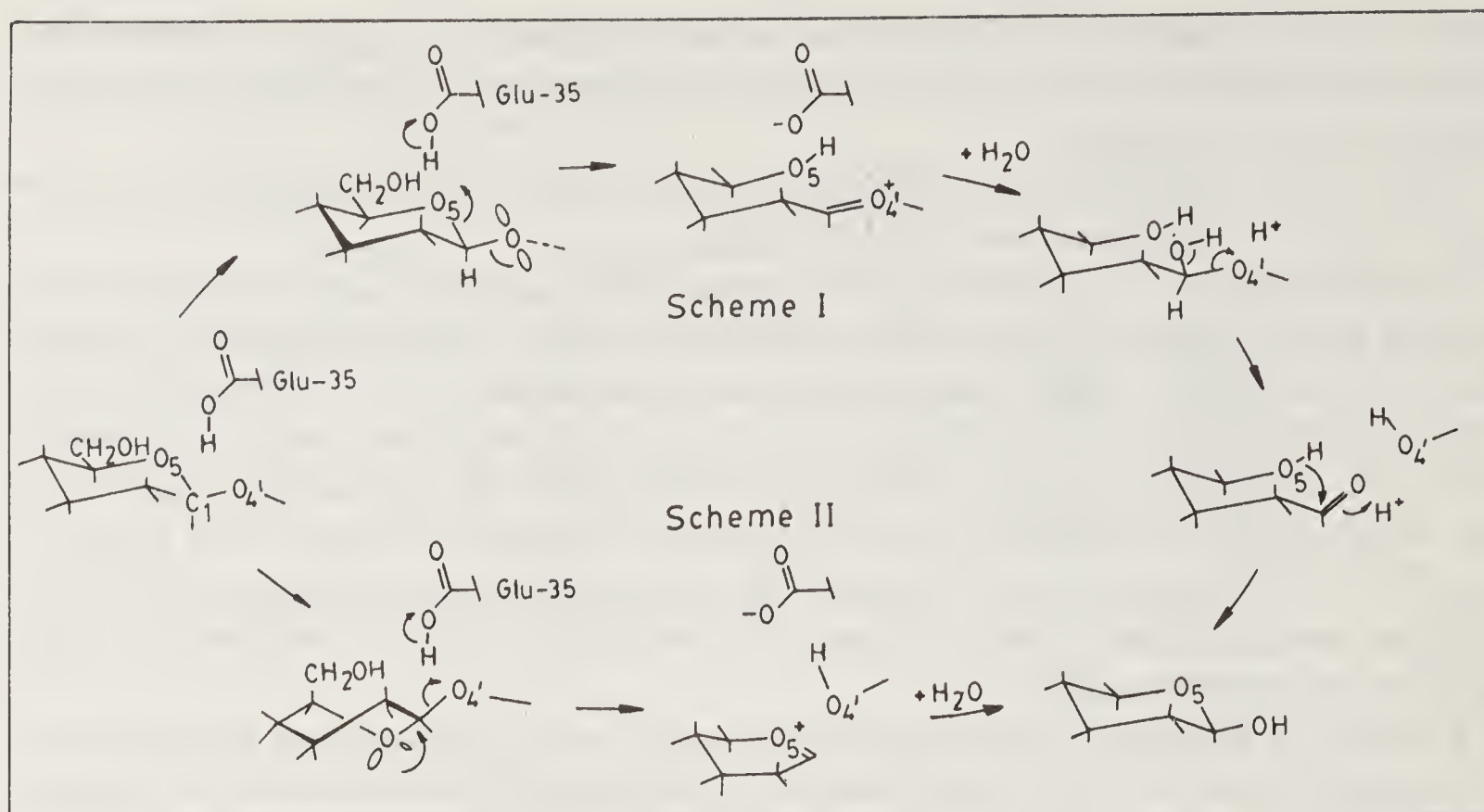


Figure 10. Proposed mechanism of action of lysozyme from molecular dynamics simulations (scheme I) (40) and X-ray crystallographic studies (37) (scheme II).

and extend the information obtained from X-ray studies in understanding molecular interactions.

References

1. Quioco, F. A. & Vyas, N. K. (1984) *Nature (London)* **310**, 381–386.
2. Makinen, M. W., Troyer, J. M., Werff, H. V. D., Berendsen, H. J. C. & van Gunsteren, W. F. (1989) *J. Mol. Biol.* **207**, 201–216.
3. Kollman, P., Wipff, G. & Singh, U. C. (1985) *J. Am. Chem. Soc.* **107**, 2212–2219.
4. Pincus, M. R. & Scheraga, H. A. (1981) *Acc. Chem. Res.* **14**, 299–306.
5. Post, C. B., Brooks, B. R., Karplus, M., Dobson, C. M., Artymiuk, P. J., Cheetham, J. C. & Phillips, D. C. (1986) *J. Mol. Biol.* **190**, 455–479.
6. Ghosh, I. & Rao, V. S. R. (1982) *Int. J. Biol. Macromol.* **4**, 130–136.
7. Ghosh, I. & Rao, V. S. R. (1984) *J. Biomol. Str. Dyn.* **2**, 29–40.
8. Sekharudu, Y. C. & Rao, V. S. R. (1984) *J. Biomol. Str. Dyn.* **2**, 41–54.
9. Vinyaka, C. R. & Rao, V. S. R. (1987) *J. Biomol. Str. Dyn.* **4**, 1095–1103.
10. Sekharudu, Y. C. & Rao, V. S. R. (1984) *Int. J. Biol. Macromol.* **6**, 337–347.
11. Biswas, M., Sekharudu, Y. C. & Rao, V. S. R. (1987) *Carbohydr. Res.* **160**, 151–170.
12. Mukhopadhyay, C. & Rao, V. S. R. (1988) *Int. J. Biol. Macromol.* **10**, 217–226.
13. Rao, V. S. R., Biswas, M., Mukhopadhyay, C. & Balaji, P. V. (1989) *J. Mol. Struct.* **194**, 203–214.
14. Rao, V. S. R., Reddy, B. V. S., Mukhopadhyay, C. & Biswas, M. (1990) in *Computer Modelling of Carbohydrate Molecules*, (eds. French, A. D. & Brady, J. W.), Chapter 22, ACS Symposium No. 430, ACS Washington, pp. 361–376.
15. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963) *J. Mol. Biol.*, **7**, 95–99.
16. Fischer, E. (1884) *Chem. Ber.*, **27**, 2985–2993.
17. Silhavy, T. J., Ferenci, T. & Boos, W. (1978) in *Bacterial Transport*, ed. Rosen, B. P. (Dekker, New York), pp. 127–169.
18. Hogg, R. W. & Englesberg, E. (1969) *J. Bacteriol.* **100**, 423–432.
19. Quioco, F. A., Gilliland, G. L. & Phillips, G. N. Jr. (1977) *J. Biol. Chem.* **252**, 5142–5149.

20. Newcomer, M. E., Gilliland, G. L. & Quijcho, F. A. (1981) *J. Biol. Chem.* **256**, 13213–13217.
21. Argos, P., Mahoney, W. C., Hermodson, M. A. & Hane, M. (1981) *J. Biol. Chem.* **256**, 4357–4361.
22. Mukhopadhyay, C. & Rao, V. S. R. (1989) *Int. J. Biol. Macromol.* **11**, 194–200.
23. Mukhopadhyay, C. (1989) PhD Thesis, Indian Institute of Science, Bangalore, India.
24. Brown, C. E. & Hogg, R. W. (1972) *J. Bacteriol.* **111**, 606–613.
25. Goldstein, I. J. & Hayes, C. E. (1978) *Adv. Carbohydr. Chem. Biochem.* **35**, 127–340.
26. Hardman, K. D. & Ainsworth, C. F. (1972) *Biochemistry* **11**, 4910–4919.
27. Hardman, K. D. & Ainsworth, C. F. (1976) *Biochemistry* **15**, 1120–1128.
28. Sekharudu, Y. C. & Rao, V. S. R. (1984) *Curr. Sci.* **53**, 403–406.
29. Sekharudu, Y. C. & Rao, V. S. R. (1985) *J. Biosci.* **8**, 389–401.
30. Sekharudu, Y. C. & Rao, V. S. R. (1985) *Int. J. Biol. Macromol.* **17**, 349–356.
31. Biswas, M., Sekharudu, Y. C. & Rao, V. S. R. (1986) *Int. J. Biol. Macromol.* **8**, 2–8.
32. Sekharudu, Y. C., Biswas, M. & Rao, V. S. R. (1986) *Int. J. Biol. Macromol.* **8**, 9–19.
33. Brewer, C. F., Sternlicht, H., Marcus, D. M. & Grollman, A. P. (1973) *Biochemistry* **12**, 4448–4457.
34. Reddy, B. V. S. & Rao, V. S. R. (Unpublished Data).
35. Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb, A. J., Dodson, E. J., Papiz, M. Z., Wan, T. & Campbell, J. (1989) *EMBO J.* **8**, 2189–2193.
36. Jolles, P. & Jolles, J. (1984) *Mol. Cell. Biochem.* **63**, 165–189.
37. Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967) *Proc. R. Soc. London, Ser. B* **B167**, 378–388.
38. Warshel, A. & Levitt, M. (1976) *J. Mol. Biol.* **103**, 227–249.
39. Pincus, M. R. & Scheraga, H. A. (1981) *Biochemistry* **20**, 3960–3965.
40. Post, C. B. & Karplus, M. (1986) *J. Am. Chem. Soc.* **108**, 1317–1319.

The Ramachandran Plot and Beyond: Conformational and Topographical Considerations in the Design of Peptides and Proteins

Victor J. Hruby[†] and Gregory V. Nikiforovich

Department of Chemistry, University of Arizona, Tucson, AZ 85721 USA.

Introduction

Efforts to understand the relationships of peptide and protein structures to their biological activities have been one of the central concerns of modern molecular biology. In the case of peptide hormones and neurotransmitters, this problem has further been exacerbated by the fact that these compounds tend to be relatively small polypeptides with a high degree of conformational flexibility at physiological temperatures in most milieu. Nonetheless, it is well known that small peptide hormones and neurotransmitters of 3 to 30 residues can bind with high affinity to their receptors and other acceptor molecules (K_D 's $\approx 10^{-8}$ to 10^{-12} M) and in some cases with high specificity. (At the same time, they are also often rapidly degraded by proteolytic enzymes, and this requires different binding interactions, though the binding affinities to enzymes generally are less – $\sim 10^{-4}$ to 10^{-6} M.) A central dogma of molecular biology includes an assumption that the biological activity of a peptide or protein is directly related to its three-dimensional structure. Thus, the specific identification of those structural features of peptide hormone or neurotransmitter ligand that were responsible for its biological activities are of central importance in understanding the molecular mechanisms involved in their bioactivities.

Initially, these efforts were primarily directed toward the functional groups that are found in peptides. Since polypeptides are made up of repeating backbone units, it was assumed that the side chain groups on the α -carbon were primarily responsible for bioactivity. Thus the aliphatic, aromatic, acidic, basic, hydrophilic and hydrophobic

[†] Author to whom reprint requests should be addressed.

Abbreviations: All amino acid residues except glycine are of the L-variety unless otherwise indicated. The abbreviations used are in accord with the recommendation of the IUPAC-IUB Commission on Biochemical Nomenclature (1972) *J. Biol. Chem.* **247**, 977–983. Other abbreviations include: pen = penicillamine (β,β -dimethylcysteine); Nle, norleucine, α -aminohexanoic acid; Tic, tetrahydroisoquinoline carboxylic acid.

side chain groups were modified to determine those groups primarily responsible for the biological activity of the peptide in *in vitro* and *in vivo* bioassays. When it became possible later to perform binding assays (radioreceptor assays) to measure specific binding to the receptor or acceptor molecules of interest, it became possible to examine the relative importance of various side chain groups to receptor or acceptor binding. An important observation of examining bioactivity from the perspectives of both bioactivity and binding was that there was often not a one-to-one correlation between the two. That is, the bioassays and the binding studies did *not* often lead to the same structure–activity relationships. These observations suggested that the structural features important for binding and biological transduction might be different, and led us to suggest, for example, that different features of structure were used for interaction of oxytocin agonist analogues and antagonist analogues with the oxytocin receptors (1, 2).

In this paper we will discuss the development of ideas related to the design of peptide hormones and neurotransmitters with special emphasis on the role the Ramachandran plot and conformational analysis has played, from our perspective, in these developments. We will conclude with a discussion of topographical consideration in the design of peptides.

Before the Ramachandran Plot

As already mentioned, the early studies of peptide structure–biological activity relationships centered on efforts to determine the importance of the side chain groups of individual residues to the bioactivity of the peptide. This classical approach involves single or multiple substitution of amino acid residues for other amino acid residues. A usually unspoken assumption of such studies is that these structural perturbations do not affect the backbone conformation significantly. However, this is not often the case, and hence such structure–activity studies can often give misleading results in the sense that the local perturbation of structure has more global effects on the overall structure and the way in which the peptide can interact with a particular receptor. Nonetheless, *systematic* studies in this area utilizing the concept of incremental structural change can provide useful, even essential, information related to receptor requirements. A few examples are given below.

Pseudoisosteric replacements of the side chain moieties can be very valuable since such structures can maintain similar steric effects, but can modify electronic and stereoelectronic effects. For example, norleucine (Nle) is pseudoisosteric to Met ($-\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_3$ vs. $-\text{CH}_2-\text{CH}_2-\text{S}-\text{CH}_3$) and to Orn ($-\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_3$ vs. $-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{NH}_2$), but in the later case the basicity of the two groups (and hence the hydrophilicity and hydrophobicity) are markedly different. Other obvious examples of this kind include:

Val for Thr ($-\text{CH}(\text{CH}_3)_2$ vs. $-\text{CH}(\text{OH})\text{CH}_3$),

Glu for Gln ($-\text{CH}_2\text{CH}_2\text{CO}_2\text{H}$ vs. $-\text{CH}_2\text{CH}_2-\text{CONH}_2$),

Asn for Leu ($-\text{CH}_2-\overset{\text{O}}{\underset{\parallel}{\text{C}}}-\text{NH}_2$ vs. $-\text{CH}_2-\underset{\text{CH}_3}{\underset{|}{\text{CH}}}-\text{CH}_3$)

and His for Phe ($\text{CH}_2-\text{CH}=\text{N}-\text{CH}=\text{N}-\text{H}$ vs. $\text{C}_6\text{H}_5-\text{CH}_2$).

Another approach has been to replace all L-amino acids one at a time by Ala or Gly, and Gly by D- and L-alanine. The replacement of each L-amino acid residue by its isomeric D-amino acid residue, one at a time, has also been used to determine the stereoisomeric requirements of each residue. Each of these approaches has obvious limitations due to the likely perturbation of the conformational properties in many cases.

Most peptide hormones and neurotransmitters have a critical aromatic amino acid that is important to binding and transduction, and hence substituted aromatic amino acid residues play an important role in the design of analogues. For example, Tyr residues can be substituted by O-alkylated or O-acylated residues, or the p-group in Phe or Tyr can be replaced systematically by a wide variety of groups (e.g. CH₃, OR, R, F, Cl, Br, I, CN, NO, NO₂, etc.) that allow one to assess the steric and stereoelectronic effects at this position. The aromatic moiety can also be replaced by bulky groups such as naphthyl and adamantyl. If a disulfide bond exists, it can be replaced by "carba," or each Cys residue by an α -aminobutyric acid (Abu), or serine by Cys(SH) to examine the importance of a ring structure to biological activity.

Though the above approaches have limitations (many other suggestions could be made), they have often provided critical leads in the development of potent, receptor selective agonist and antagonist analogues.

Considerations at the Level of the Ramachandran Plot

Once structure–function studies on peptides had been initiated, it was soon realized that conformational properties must be important. The introduction of the Ramachandran plot (3, 4) has had a dramatic and critical effect on these considerations. The Ramachandran plot represents the intramolecular atom–atomic interactions within a dipeptide unit backbone as a function of two dihedral angles, namely ϕ and ψ , which are the angles of rotation around the adjacent valence bonds N—C ^{α} and C ^{α} —C', respectively. This revolutionary idea has focused the problem of describing nearest-neighboring interactions in a peptide by reducing the number of variables from the tens of Cartesian atomic coordinates to these two dihedral angles only. In this sense, it has provided a method of reducing a problem in multi-dimensional space to that of examining only a few essential variables that are present such that the problem can be solved just by testing all possible combinations of values for the essential variables. Obviously, to make the proper choice of variables for this kind of mathematical approach, one must carefully consider assumptions that derive from the physical nature of the problem. In the case of the Ramachandran plot, these assumptions were the following:

- i) The rigidity of valence geometry, i.e. of the bond distances and the bond angle values;
- ii) The planar *trans* conformation of the peptide bond; and
- iii) The representation of atomic interactions by hard-sphere potentials.

The first of these assumptions appeared to be very suitable for the overwhelming majority of conformational problems in peptides and proteins. In this regard, it is interesting to note that the distribution of the points representing X-ray data on the (ϕ , ψ) values for single residues in proteins is generally more consistent with the

Ramachandran plot than with other potential plots which consider non-rigid valence geometry (see, e.g. 5). The assumption of the peptide bond planarity can lack validity in the case of small cyclic compounds (see, e.g. X-ray data on some peptides (6)). The *cis* peptide bond sometimes occurs in the same cyclic compounds and in N-substituted amino acid residues including Pro. The hard sphere potential field describing atom–atomic interactions is the most simplified assumption used in the original Ramachandran plot. It was clear from the beginning that the conformational states which are “allowed” by the original Ramachandran plot were not very accurate in their conformational energies, and in the past several years, several more realistic potential fields have been proposed by different authors (see reviews in 7). In any case, throughout these developments, the Ramachandran plot has served as an important guide in the design of biological active peptides using conformational considerations, and it is to these considerations that we now turn.

Design at the Level of the Ramachandran Plot

Despite the limitations inherent in the Ramachandran plot, its general usefulness in peptide design is excellent. Normally in a design process, the investigator proposes a possible model for a “biologically active conformation,” i.e. the conformation or specific conformational properties that are proposed to fit the complementary surface of the specific receptor providing the binding and specificity of the peptide–receptor interaction. Some elements of this model can be deduced using the Ramachandran plot to evaluate directly the structure–activity relationships in a series of the analogues for a given peptide. The Ramachandran plot can be very roughly divided into four quadrants, namely the upper left (corresponding to $\phi < 0^\circ, \psi > 0^\circ$), the lower left ($\phi < 0^\circ, \psi < 0^\circ$), the upper right ($\phi > 0^\circ, \psi > 0^\circ$) and the lower right ($\phi > 0^\circ, \psi < 0^\circ$). These quadrants have been denoted by some authors as B, R, L and H, respectively. The B quadrant contains the (ϕ, ψ) points corresponding to a β -structure and the R and L quadrants contain points corresponding to the right and the left handed α -helices, respectively. The amino acid residues involved in a peptide chain can be divided into four groups according to their Ramachandran plots. Those groups are L-amino acid residues, D-amino acid residues, Pro and Gly. (Note that under the assumption of rigid valence geometry the C^β -atom is rigidly attached to the asymmetric C^α -atom and can therefore be regarded as a part of the peptide backbone. This is the reason for the differences in the Ramachandran plots between L- and D-amino acid residues.) At the same rough “quadrant” approximation, the B, R, and L quadrants can be considered as the sterically allowed and the H quadrant as forbidden for L-amino acid residues. On the other hand, the H, L, and R quadrants are allowed, and the B quadrant is forbidden for a D-amino acid residue (i.e., the Ramachandran plots for the L- and D-amino acid residues are symmetrical as to the rotation by 180° around the axis at the center of the plot). All four quadrants B, R, L, and H are allowed for a Gly residue and only two of them, namely, B and R are allowed for a Pro residue (where the ϕ value is fixed at -60°).

The effect of a single amino acid substitution on the biological activity of an analogue can be evaluated in terms of the possible backbone conformations for a substituted residue. (Note that this implies another assumption, namely that the changes in biological activity of the analogue can be assigned to conformational changes in the

Table 1. Possible backbone conformations that can be deduced from structure–activity relationships in the case of single amino acid residue substitutions. The suggestions when the resulting analogue is active are above the horizontal bars, and below the bars when the resulting analogue is inactive.

	Substituted for:		
	L-Amino Acid	D-Amino Acid	Pro
L-Amino acid	—	R/L — B	B/R — L
D-Amino acid	R/L — H	—	R — H/L
Pro	B/R — ?	R — B	—
Gly	not H — H	not B — B	B/R — H/L

backbone only. It goes without saying that such an assumption has limited use.) For instance, when a residue possessing a more sterically allowed Ramachandran plot profile (like Gly) is substituted for another one possessing a more restricted Ramachandran plot profile (like Pro), and the resulting analogue is still active, it is very likely that the biologically active backbone conformer of the parent peptide in this particular position of peptide chain is characterized by sterically allowed quadrants which overlap in both plots (B/R quadrants in this case). And, *vice versa*, when the resulting analogue is inactive, it might be indicative of the existence of non-overlapping allowed quadrants in the analogue and the parent peptide (like H/L in the example considered). Interpretations of this kind are summarized in table 1. The suggestions for possible backbone conformations are given above the horizontal bars when the resulting analogue is active, and below when the analogue is inactive. These kinds of considerations have been applied, for instance, in suggestions concerning the possible bioactive conformers of bradykinin and enkephalin. In both cases, substitution of a Gly residue (Gly 4 in bradykinin and Gly 2 in enkephalin) by a D-Ala residue resulted in an active analogue, and by an L-Ala in an inactive analogue (see, e.g. 8 and 9, respectively). These findings were interpreted as indicating that the backbone conformations of the Gly residues in the corresponding bioactive conformers were not extended ones (like those in B quadrant). Rather it could be suggested that the Gly residues in question can serve as “conformational hinges” leading to folding of the backbones (10, 11). Evidence for such folding was obtained by spectroscopical measurements (12, 13) and by synthesis of active cyclic analogues of both bradykinin (14) and enkephalin (15).

The presence of Pro (and any other N-Me-amino acid residue) in a peptide chain causes an additional effect in the Ramachandran plot for the *preceding residue*. In this

case, the R quadrant appears to be sterically forbidden for an L-amino acid residue, and the L quadrant for a D-amino acid residue. (That's why Pro is known as an α -helix breaker.) This approach can be utilized to investigate the backbone conformation of the X residue in an X-Pro (or X-N-Me-residue) fragment for active or inactive analogues. These considerations also have been utilized to suggest aspects of biologically active backbone conformers of tetragastrin Met 2 residue), substance P (Phe 8–Gly 9 fragment), bradykinin (conformer related to B₂ receptors), and angiotensin (Arg 2 and Phe 8 residues). This approach can be especially useful when applied to select possible biologically active backbone conformers from a set of low-energy conformations for a given peptide obtained independently by energy calculations (for a review see ref. 10).

**Other Uses of the Ramachandran Plot Especially
in Examination of β -Turns**

The considerations discussed above can also be applied in the design of peptide conformations involving backbone structures consisting of two or more neighboring residues. Structures such as α -helices, β -sheets, extended structures, etc. are evident at the individual amino acid residue, and can be extended to n residues by a similar analysis. Particularly instructive is the utilization of the Ramachandran plot in the examination of peptide chain reversal structures, now commonly known as β -turns. Indeed, the possibility for these structures though previously recognized by experimentalists (for a review of early literature see ref. 16) was first explicitly defined utilizing the theoretical approach of Ramachandran *et al.* (17, 18). The importance of this kind of backbone structure in peptide and proteins was quickly recognized and exploited by many investigators of peptide and protein structure and function, and

Table 2. Types of β -turns (from [20])

Type	φ_{i+1} (°)	ψ_{i+1} (°)	φ_{i+2} (°)	ψ_{i+2} (°)
I	− 60	− 30	− 90	0
I'	60	30	90	0
II	− 60	120	80	0
II'	60	− 120	− 80	0
III	− 60	− 30	− 60	− 30
III'	60	30	60	30
IV	any of bend types I through III' with two or more angles φ_{i+1} , ψ_{i+1} , φ_{i+2} or ψ_{i+2} differing by at least 40° from typical angles given above.			
V	− 80	80	80	− 80
V'	80	− 80	− 80	80
VI	a <i>cis</i> proline at position $i + 2$			
VII	a kink in the protein chain created by $ \psi_{i+1} \approx 180^\circ$ and $ \varphi_{i+2} < 60^\circ$ or $ \psi_{i+1} < 60^\circ$ and $ \varphi_{i+2} \approx 180^\circ$.			

exhaustive reviews have been written (16, 19). Table 2 lists the various types of β -turns as they are more formally classified (20). It is evident that the conformations of the $i + 1$ residue in β -II'-turn and the $i + 2$ residue in β -I', β -II- and β -V-turns are found mainly with H quadrant of the Ramachandran plot. Hence these particular types of β -turns should exist in peptides and proteins mainly with Gly in corresponding positions, and perhaps can be additionally stabilized by the D-amino acid substitutions (16, 20). Other immediate observations are that destabilization of β -turns can be achieved for instance by Pro substitution for $i + 2$ residue in β -I- and β -III-turns. Thus, in principle, one can control the conformation of a two-residue fragment taking into account the Ramachandran plot for a single residue only.

It was immediately obvious that this structural feature was important to a variety of cyclic peptides, and many investigators examined or re-examined cyclic peptide antibiotics, ionophores, etc. with these structures in mind using X-ray analysis, nuclear magnetic resonance spectroscopy, etc. (for reviews see refs. 16, 19, 21, 22). In the peptide hormone area, the ideas were applied to oxytocin agonists (23, 24) and antagonists (1) which were proposed to have reverse turn structures based on a variety of experimental studies (e.g. 25). From the standpoint of peptide design, the stabilization of the reverse turn conformation in the cyclic peptide hormone somatostatin (26), and the stabilization of the reverse turn structure in the linear peptide hormone α -melanotropin (27) leading to the design of a superpotent cyclic structure (28) are particularly instructive. In the former case, the cyclic 14-amino acid peptide somatostatin, H-Ala-Gly-Cys-Lys-Asn 5-Phe-Phe-Trp-Lys-Thr 10-Phe-Thr-Ser-Cys-OH was systematically examined for those features of structure and conformation important to activity at somatostatin receptors. A β -turn at -Phe 7-Trp 8-Lys 9-Thr 10- was proposed to be important, and its stabilization by substitution of a D-Trp 8 for L-Trp 8 gave a more potent analogue consistent with this suggestion. Eventually it was possible to prepare cyclic hexapeptides such as *c*-(Pro-Phe-D-Trp-Lys-Thr-Phe) which were as potent or even more potent than the original hormone (see 29 for a review).

The design consideration that went into the conversion of α -melanotropin, Ac-Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val-NH₂, a linear 13-amino acid peptide, to a cyclic conformationally constrained β -turn structure were somewhat different. The first step, which involved the stabilization of a proposed β -turn conformation by the substitution of the Phe 7 residue by D-Phe 7 to give Ac-[Nle 4, D-Phe 7] α -MSH (the Met 4 \rightarrow Nle 4 substitution was an isosteric substitution, made to eliminate the sulfur atom which was easily oxidized with loss of potency), provided a superpotent, prolonged acting, superagonist (27). This led to the concept of folding of the backbone structure on interaction with the receptor (quasicyclization) as an important conformational effect that would be critical to the biological activity of linear peptide hormones and neurotransmitters. This concept was further tested by the synthesis of the pseudoisosteric cyclic analogue of α -melanotropin, Ac-[Cys 4, Cys 10] α -MSH, which was found to be superpotent at its minimum effective dose in the frog skin assay, consistent with the proposed bioactive conformation for melanotropin (28). Investigations in several laboratories have demonstrated the power of the concepts of conformational stabilization by β -turns and subsequent cyclization for further conformational stabilization in the design of biologically active peptide hormones and neurotransmitters (for a review, see ref. 30).

Conformational Considerations with More Explicit Inclusion of Side Chain Moieties in Peptide and Peptidomimetic Design

Though it utilized considerations from the Ramachandran plot in its design, the above discussed conversion of the linear peptide hormone α -melanotropin into cyclic analogues also depended on design considerations beyond the plot, namely the properties and roles of side chain groups in peptide conformation and their relation to biological activity. The Ramachandran plot relates to the interactions inside the dipeptide unit only. This implies that there are no significant differences between the Ramachandran plots that describe amino acid residues possessing quite different side chain groups. However, a more realistic view of peptide conformations suggests that intramolecular interactions in which side chain groups are involved should be taken into account as well. Thus, for example, the electrostatic interactions that are possible between positively and negatively charged functional groups would seem to be of great potential importance. For example, such interactions might help stabilize folded conformations that can accommodate salt bridge formation between oppositely charged groups that are in close proximity as a result of a particular folded structure. This was an important consideration in the design of cyclic lactam analogues of α -melanotropin (31, 32). These design studies incorporated the β -turn concept in conjunction with considerations from molecular dynamics studies, and led to the design of superpotent, prolonged acting cyclic lactam analogues e.g. $\text{Ac-Nle 4-Asp-His-D-Phe-Arg-Trp-Lys-NH}_2$ (31, 32) which are half the size of α -melanotropin. This also has been suggested, for example by energy calculations and spectroscopic data for the interaction of the guanidine group of Arg 1 and the C-terminal carboxyl in bradykinin, $\text{Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg}$ (12). The idea of the possibility of salt bridge formation in bradykinin in water solution has been criticized (33). It was pointed out that electrostatic interactions in water solution should be rather weak due to hydration of the interacting charged groups. Nonetheless, it is interesting that the cyclic bradykinin analogues in which the assumed salt bridge was replaced by a covalent bond showed the same potency in biological assays as bradykinin itself, with an additional feature of prolonged action (14), as seen for instance in $\text{Lys-Pro-Pro-Gly-Phe-Gly-Phe-Arg}$. It has been proposed that dehydration of charged groups during the process of receptor binding provides an explanation for the observed activity of certain cyclic tuftsin analogues (34) e.g. Thr-Lys-Pro-Arg . There is also some experimental evidence for the existence of quasicyclic peptide conformers stabilized by salt bridges in solution for angiotensin (35) and a few model compounds (36). These and other findings provide a conceptual basis for the design of cyclic analogues of linear peptides in various peptide series, utilizing the additional concept of pseudoisosteric cyclization (30).

Some additional instructive examples can be found in the design of cyclic enkephalin analogues with a high potency and selectivity for mu or for delta opioid receptors. The enkephalins, $\text{H-Tyr-Gly-Gly-Phe-Leu-OH}$ ($[\text{Leu}^5]$ enkephalin) and $\text{H-Tyr-Gly-Gly-Phe-Met-OH}$ ($[\text{Met}^5]$ enkephalin) are linear, conformationally flexible pentapeptides, believed to be the natural endogenous ligands for opioid receptors. Their high degree of conformational flexibility, which enables them to assume scores of conformations at physiological temperatures, has made it difficult to determine their "bioactive conformation." Though numerous suggestions were made that a folded conformation was most likely at the opioid receptor, these suggestions were never

tested by critical experiments involving the design of analogues to test these hypotheses until Schiller and co-workers (37, 38) found that substitution of an ϵ -amino-D-amino acid in position 2 followed by lactam cyclization to the C-terminal carboxyl group gave cyclic analogues of [Leu⁵]enkephalin such as H-Tyr-D-Lys-Gly-Phe-Leu, which were as potent or more so than [Leu⁵]enkephalins. Perhaps equally important was the observation that the linear compounds were non-receptor selective, while the cyclic ones were quite μ opioid receptor selective. More recently, design of cyclic enkephalin via side chain to side chain cyclization led to analogues such as Tyr-D-Orn-Phe-Asp-CONH₂ which are highly potent mu opioid receptor ligands (39). In another approach, Mosberg, Hruby and co-workers utilized the idea of pseudoisosteric cyclization (quasicyclization involving the Met 5 side chain and pro-*R* proton of Gly 2) in conjunction with structural modifications to further *decrease* conformational flexibility via the geminal dimethyl effect using a D-Pen (D- β - β -dimethylcysteine) residue in the design of highly δ opioid selective analogues such as [D-Pen², D-Pen⁵]enkephalin, DPDPE) (40, 41). The 14-membered disulfide-containing ring possesses two β , β -dimethylsubstituents in the ring structure, and the analogue shows very high δ opioid receptor selectivity (41). Subsequent conformational studies of these and related analogues utilizing NMR and conformational energy calculations (e.g. 42, 43) have demonstrated that these compounds have restricted backbone conformations, and that the geminal dimethyl groups, especially those in the 2 position, are of critical importance to the conformational properties of these compounds and to their receptor selectivity.

The studies reported above for somatostatin, melanotropin, angiotensin, bradykinin and enkephalin analogues all demonstrate in different ways that whereas the backbone conformations are important to biological action, the proper consideration of side chain groups, their steric and stereoelectronic properties, must also be considered in designing peptide analogues for interactions with biologically relevant acceptor molecules (receptors, enzymes, antibodies, etc.). In other words, the backbone conformation alone (in the sense of the Ramachandran plot for example) is more often than not a sufficient consideration in the design of highly potent and specific peptide ligands.

Considerations Related to the Ramachandran Plot in Peptidomimetics

In the context of this discussion of the significance of side chain moieties in peptide design, an additional concept deserves some discussion. This concept has particular current relevance because of the current interest in developing peptidomimetics. This is the idea that for many biologically active peptides, the peptide backbone (i.e. -NH- α CH-CO-) serves primarily as a "carrier" of functionally important side chain groups, that is, the primary role of the peptide backbone is to arrange the proper stereo-orientation of certain side chain functional groups important for a specific receptor or acceptor molecule. Alternatively of course, it may be imagined that the primary role for a specific backbone conformation (say a β -turn) is to properly orient amino acid residues to the N-terminal and C-terminal of the specific conformational feature within the peptide. In any case, a peptide mimic could serve in the stead of the "real" peptide. In the latter case, conformational analysis at the level of the Ramachandran plot should

be extremely valuable. In the former case, conformational analysis at the level of the Ramachandran plot, probably at best would be incomplete and at worst misleading.

One approach that has been under development in many laboratories has been that of the amide bond replacement (see ref. 44 for a review). It is not clear at this point, even at the level of the Ramachandran plot, just which amide bond replacement moieties are truly isosteric replacements, i.e. consistent with the Ramachandran plot. However a number of investigations have been undertaken to examine this question. For example, Goodman and co-workers have been developing the use of retroinverso peptides (45) as a useful starting point for the design of peptidomimetics, and have used the cyclic enkephalins of Schiller as a suitable peptide to examine for these purposes (e.g. 46). Such peptides and other peptide bond modified analogues hold considerable promise as peptidomimetics, and thus far have been very useful especially in the design of proteolytic enzyme inhibitors. Another area where there has been considerable progress has been in the design of mimics of peptide backbone structures like α -helices, β -turns, and β -sheets (see for example 47–51). The design of peptide mimetics for the receptor binding site of a peptide hormone or neurotransmitter has had only a modest success to date, probably because most such mimetics have not possessed appropriately positioned side chain groups. However, a few recent successes bear mentioning because they suggest that more important successes may soon appear in this area. For example, Hirschmann and co-workers (52) have designed a mimetic based on D-glucopyranoside as a scaffold which is intended to mimic the key side chains in somatostatin, namely the Trp and the Lys side chain in the $i + 1$ and $i + 2$ positions of the β -turn (-Phe-D-Trp-Lys-Thr-) in potent cyclic somatostatin analogues (see above). The compound designed was biologically active, albeit 300–400 times less potent than somatostatin. In yet another approach, Olson and co-workers (53, 54) have designed β -turn mimetics and a peptidomimetic of thyrotropin releasing hormone TRH (p Glu-His-Pro). In the latter case, the mimetic was designed to mimic the side chain groups in TRH using a cyclohexane-like backbone. The analogue was found to be quite potent (54).

Design Based on the Topographical Features of Peptide Hormones and Neurotransmitters

Even in highly constrained cyclic polypeptides, there is often considerable conformational flexibility because the side chain moieties may have considerable conformational freedom by rotation about the χ_1 , χ_2 and other side chain torsional angles. Careful evaluation of the preferred side chain conformations will be a critical element in the biological activities of the peptide and in its topographic properties. Ultimately, rational design requires that topographical design be an integral part of any peptide or protein design approach. Topography in the context discussed here is concerned with the relative, cooperative three-dimensional arrangements of the side chain groups of a polypeptide. Thus, within the context of a single specific backbone conformation, it is possible to have numerous topographical structures even for a relatively small peptide. We believe that development of this approach is a critical missing link in efforts to understand the common problems of peptide and protein design including protein folding, peptide receptor interactions, enzyme–substrate specificity, antigen–antibody

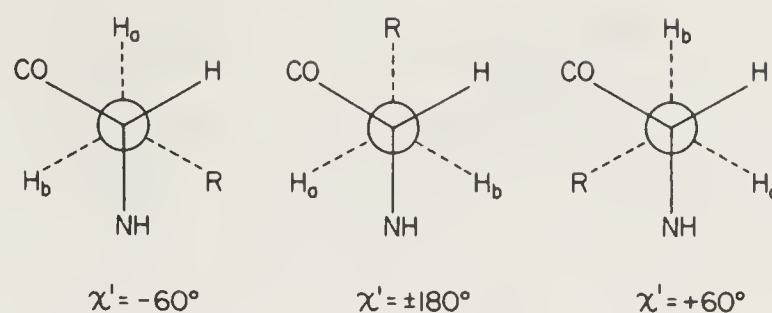


Figure 1. Newman projection of the χ'_1 side chain conformations *gauche* (–) (– 60°), *trans* ($\pm 180^\circ$) and *gauche* (+) (+ 60°) in an amino acid residue.

complementarity, etc. This is a largely unexplored area of investigation. However, we have made a few initial efforts in this direction that have been surprisingly successful.

We are seeking ways to develop this approach by appropriately fixing, constraining, or biasing the side chain conformations of particular amino acid residues to their classical low energy side chain rotamers, i.e. *gauche* (–) ($\chi_1 = -60^\circ$), *trans* ($\chi_1 = \pm 180^\circ$) and *gauche* (+) ($\chi_1 = +60^\circ$) (figure 1). In this way, we ultimately hope to be in a position to modify the *surface architecture* of a peptide to provide specific topographical relationships between a variety of side chain groups.

We are taking several approaches to do this, and will discuss a few here. One such approach is to enforce specific side chain conformations on amino acid residues. In the case of aromatic residues, this can be done, for example, by “connecting” an α -amino group to the 2' position in the aromatic ring of phenylalanine, by a methylene bridge (see figure 2). Such an amino acid residue, 1,2,3,4-tetrahydroisoquinoline-2-carboxylic acid (Tic, figure 2) can be thought of as a topographically constrained analogue of Phe. Note that it can be viewed as either an analogue of N-methylphenylalanine (disconnection at the 2'-aromatic position), or as an analogue of 2'-methylphenylalanine (disconnection at the N^α position). In the former case, the major difference in conformation is that, whereas in N-methylphenylalanine, the favored χ_1 angle would likely be *trans*, in the case of Tic, the *trans* conformation is not possible and, in fact, one would expect that only the *gauche* (–) or *gauche* (+) conformations (figure 3) would be allowed. Indeed, subsequent NMR studies have shown that this expectation is realized. A particularly interesting application of this amino acid residue has been in the investigation of the topographical structural requirements for the mu (μ) opioid receptor. We have discovered (55) that somatostatin, a cyclic 14-amino acid peptide

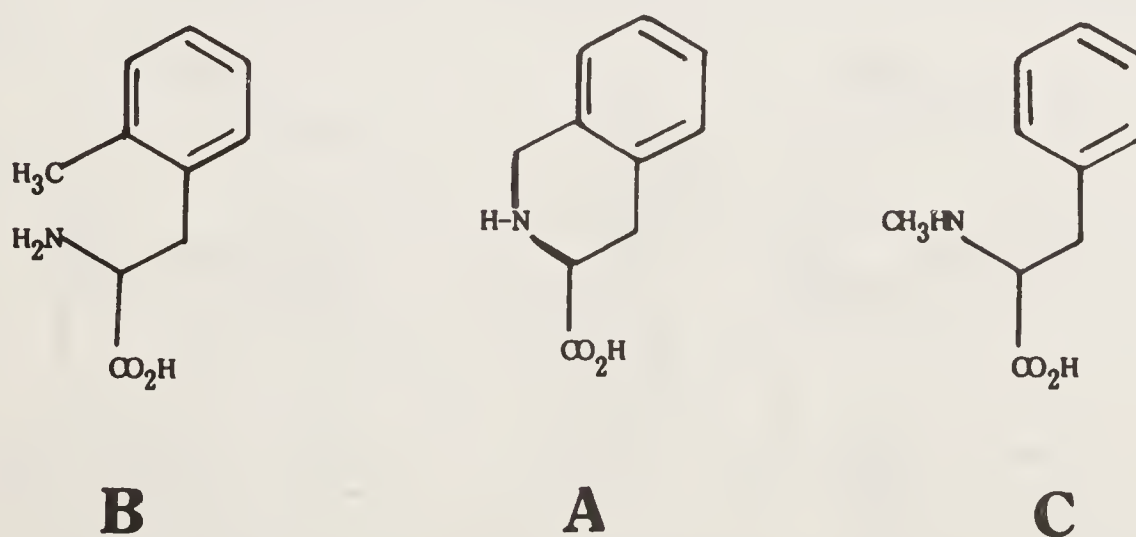


Figure 2. Structures of tetrahydroisoquinoline carboxylic acid (Tic, A), 2'-methylphenylalanine (B) and N-methylphenylalanine (C).

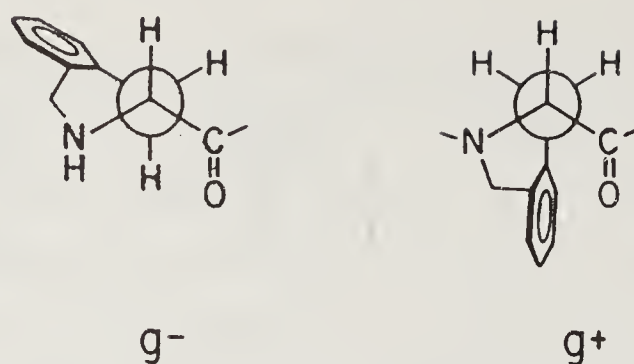


Figure 3. *Gauche*($-$) and *gauche*($+$) conformations of a tetrahydroisoquinoline carboxylate residue.

(*vide supra*), can be converted into a highly potent and selective opioid peptide with greatly reduced or essentially no bioactivity at the standard somatostatin-like receptors. Our lead compound for these purposes was the conformationally restricted peptide (CTP) $\text{D-Phe-Cys-Tyr-D-Trp-Lys-Thr-Pen-Thr-NH}_2$ (55, 56). When a D-Tic residue is substituted into the 1 position to give e.g., $\text{D-Tic-Cys-Tyr-D-Trp-Arg-Thr-Pen-Thr-NH}_2$, (TCTAP) an analogue of greater potency and selectivity for the μ opioid receptor and with reduced potency at somatostatin receptors is obtained (57). Conformational analysis (58, 59) using NMR, and molecular mechanics and molecular dynamics calculations have shown that the overall backbone conformation of these peptides is not changed by the substitution of the D-Tic residue into the 1-position. On the other hand, the side chain group is now exclusively ($> 98\%$) in the *gauche*($-$) conformation. However, when the side chain conformation in position 1 is exclusively in the *gauche*($+$) conformation as in $\text{Gly-D-Tic-Cys-Tyr-D-Trp-Arg-Thr-Pen-Thr-NH}_2$, Gly 0-(TCTAP), the compound loses about 250 fold in potency and even more in selectivity at the μ opioid receptor. In the same manner, when N-Me-D-Phe is placed in the 1 position (this biases the side chain conformation of the χ_1 torsional angle to *trans* ($\pm 180^\circ$)), the peptide also loses about 250 fold in potency and selectivity (57). These studies provide very strong evidence that topographical features alone can dramatically change both biological potency *and* selectivity without any change in

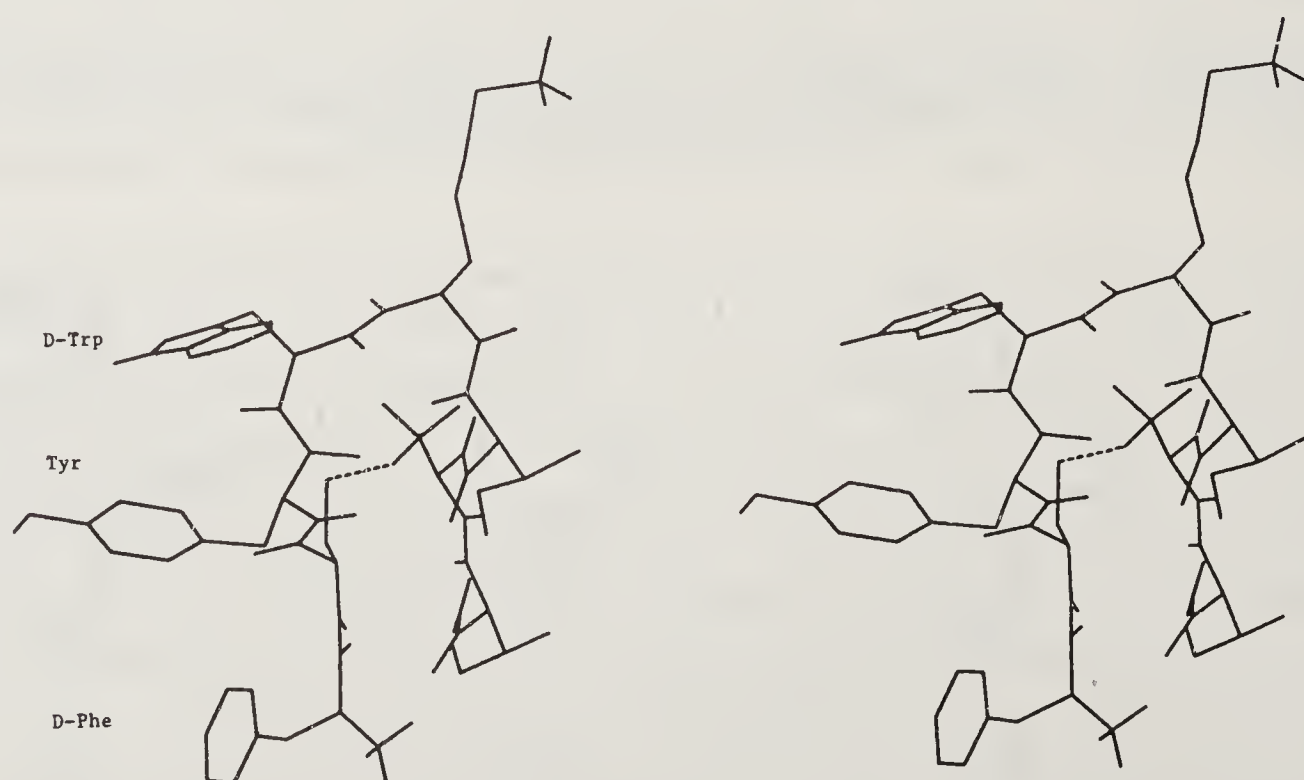


Figure 4. Stereoview of the conformation of $\text{D-Phe-Cys-Tyr-D-Trp-Lys-Thr-Pen-Thr-NH}_2$ (CTP).

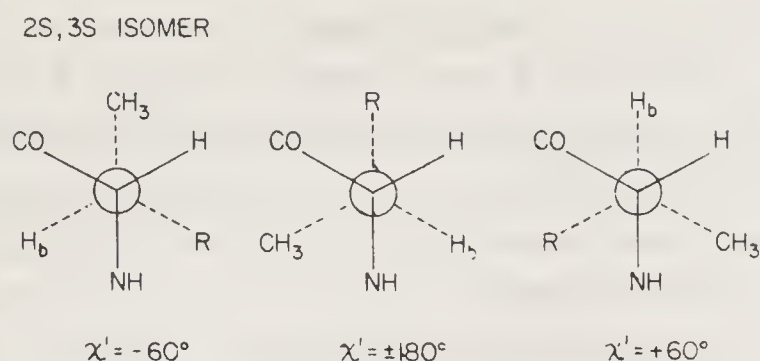


Figure 5. Newman projection of the *gauche*($-$), *trans* and *gauche*($+$) conformations of the 2S, 3S isomer of a β -methyl substituted amino acid such as β -methylphenylalanine ($R = C_6H_5$).

conformation, since extensive NMR and other biophysical studies, including molecular mechanics and molecular dynamic calculations, demonstrate that the backbone conformation is not changed by these modifications. Indeed, these studies suggest that the μ opioid receptor prefers the *gauche*($-$) conformation of the D-Phe 1 for the peptide for receptor recognition (all of these peptides are potent *in vitro* and *in vivo* μ opioid receptor antagonists). Furthermore, conformational studies have indicated that the μ opioid receptor favors a topographical arrangement of aromatic side chain groups D-Phe 1, Tyr 3 and D-Trp 4) all on the same topographical surface, but somewhat separated (in space) from one another (figure 4).

Another approach in this area is to utilize diastereotopic positions (e.g. the pro *R* and pro *S* β -hydrogens, the 2' and 6' aromatic positions, etc.) in amino acid residues for substitution into conformationally constrained analogues which have a reasonably small family of conformational possibilities. As an example, the side chain conformation of an aromatic amino acid such as phenylalanine can be biased by appropriate non-bonding interactions of vicinal substituents. For example, in L- β -methylphenylalanine, the 3S isomer (*S,S* isomer) favors a side chain conformation that is *g*($-$) (not *trans* or *g*($+$), figure 5), while the 3R isomer (*S,R* isomer) favors a side chain conformation this is *trans* (not *g*($-$) or *g*($+$), figure 6). Similar arguments could be made for the D-amino acid analogues (i.e. the *R,S* and *R,R* diastereoisomers). Indeed, quite dramatic differences in biological potency and selectivity have been seen for analogues of peptides hormones and neurotransmitters with such substitutions in our laboratory. For example, using the conformationally constrained highly delta opioid receptor agonist [D-Pen², D-Pen⁵]enkephalin (H-Tyr-D-Pen-Gly-Phe-D-Pen,

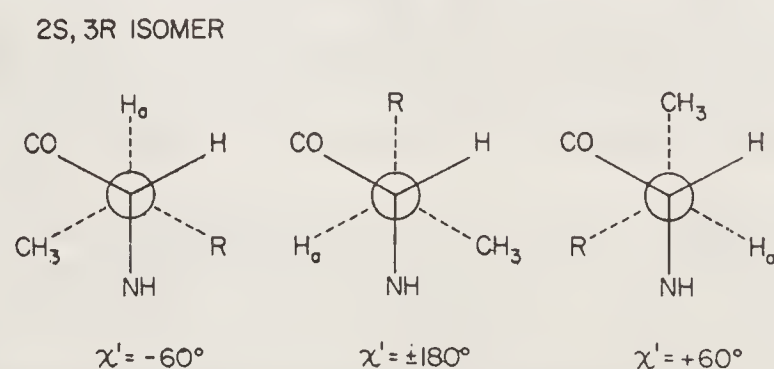


Figure 6. Newman projection of the *gauche*($-$), *trans* and *gauche*($+$) conformations of the 2S, 3R isomer of a β -methyl substituted amino acid such as β -methylphenylalanine ($R = C_6H_5$).

DPDPE), we have found by examining all diastereoisomers in which the Phe 4 residue is replaced either by β -Me-Phe 4 or β -Me-*p*-NO₂Phe 4 (60), that tremendous differences in potency (10–10,000) and in the selectivity (10–100,000) of the analogues are obtained (Hruby, Toth, Gehrig, *et al.*, unpublished). Again, NMR studies and molecular mechanics calculations indicate that the overall backbone conformations are very similar and yet their biological activities are greatly modified. These results again demonstrate that topographical considerations can be a powerful tool for the design of highly selective and potent peptide analogues. In this regard, the use of molecular mechanics calculation in conjunction with molecular modelling can provide additional insights into possible “receptor bound” conformations. A recent example of the design of receptor-bound conformer based upon topographical considerations has been undertaken in our laboratory (61). In this work, we have compared the mutual space arrangement of the α -amino group and the aromatic side chains of Tyr and Phe residues for low-energy conformers in several δ -selective opioid peptides. Among the peptides in question were both linear and cyclic ones, and they also contained the Phe residue in positions 3 and 4 of the peptide chain. It appeared that one cannot elucidate the backbone conformer(s) that would be geometrically similar for all δ -selective peptides in question, but not for μ -selective analogues when comparing the backbone structures only; it could be done by taking into account the spatial arrangements of side chains. Examples of conformers regarded as most possible δ -receptor-bound structures are depicted in figure 7.

The application of this approach to other Tic analogues, and to β , 2', 3' and other substituted amino acids, dehydroamino acids, heterocyclic amino acids, etc. for topographical control of peptide conformation provides a new, systematic approach to the design of peptide ligands.

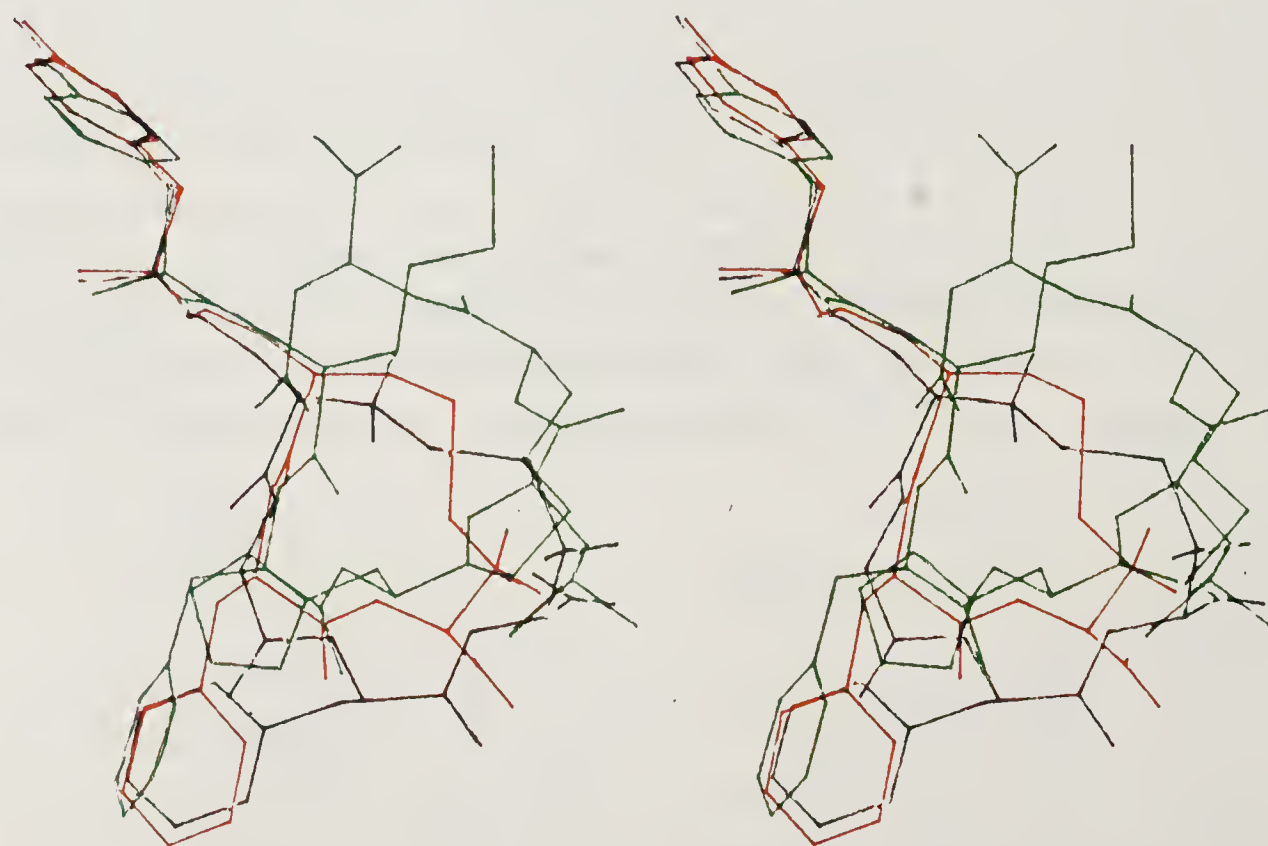


Figure 7. Stereoview of a model of the δ -opioid-receptor-bound conformations for [D-in, Pen², D-Pen⁵]enkephalin (DPDPE, red line), for Tyr-D-Cys-Phe-D-Cys (DCFPE, black line) and for Tyr-D-Met-Phe-His-Leu-Met-Asp-NH₂ (DREK, dermenkephalin, green line). All hydrogens are omitted for clarity.

Conclusions

The use of conformational considerations, especially when complemented by topographical considerations, provides a powerful tool for the rational design of peptide and protein ligands. This approach can be further complemented by application of stereoelectronic and dynamic ideas. It is to be expected that these considerations in conjunction with modern molecular mechanics and molecular dynamics calculations and computer-assisted molecular modelling will provide powerful new approaches to the design of peptides and proteins. It is exciting to consider how much progress has been made during the past twenty-five years. The importance and significance of the Ramachandran plot in these developments cannot be overstated. Nonetheless, it is clear that many additional considerations are now possible. Given the tremendous significance of peptides and proteins in living systems, it is expected that important new advances will continue well into the future.

Acknowledgements

This work was supported by grants from the U.S. Public Health Service and the National Science Foundation. We also wish to thank our many outstanding co-workers whose names are mentioned in the references for their many ideas and experimental contributions.

This paper is dedicated to Professor G. N. Ramachandran for his seminal contributions to the study of biopolymers.

References

1. Meraldi, J.-P., Hruby, V. J. & Brewster, A. I. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1373–1377.
2. Hruby, V. J. (1981) in *Topics in Molecular Pharmacology*, eds. Burgen, A. S. V. & Roberts, G. C. K., (Elsevier/North Holland Biomed. Press, Amsterdam), Vol. 1, pp. 99–126.
3. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963) *J. Mol. Biol.* **7**, 95–99.
4. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283–437.
5. Efimov, A. V. (1986) *Mol. Biol.* **20**, 208–216.
6. Karle, I. L. (1981) in *The Peptides. Analysis, Structure, Biology*, eds. Gross, E. & Meienhofer, J., (Academic Press, N.Y.), Vol. 4, pp. 1–54.
7. Hagler, A. T. (1985) in *The Peptides. Analysis, Structure, Biology*, ed. Hruby, V. J., (Academic Press, Orlando), Vol. 7, pp. 213–299.
8. Regoli, D. & Barabe, J. (1980) *Pharmacol. Rev.* **32**, 1–46.
9. Hansen, P. E. & Morgan, B. A. (1984) in *The Peptides. Analysis, Structure, Biology*, eds. Udenfriend, S. & Meienhofer, J., (Academic Press, Orlando), Vol. 6, pp. 269–321.
10. Nikiforovich, G. V. (1986) *J. Mol. Struct. Theochem.* **134**, 325–340.
11. Balodis, Yu. Yu., Nikiforovich, G. V., Grinsteine, I. V., Vegner, R. E., Chipens, G. I. (1978) *FEBS Lett.* **86**, 239–242.
12. Ivanov, V. T., Filatova, M. P., Reissmann, Z., Reutova, T. O., Efremov, E. S., Pashkov, V. S., Galaktionov, S. G., Grigoryan, G. L. & Ovchinnikov, Yu. A. (1975) in *Peptides: Chemistry, Structure and Biology*, eds. Walter, R. & Meienhofer, J. (Ann Arbor Science Publ., Ann Arbor), pp. 151–157.
13. Schiller, P. W. (1984) in *The Peptides. Analysis, Structure, Biology*, eds. Udenfriend, S. & Meienhofer, J. (Academic Press, Orlando), Vol. 6, pp. 219–268.
14. Chipens, G., Mutulis, F. & Galaktionov, S. (1980) in *Frontiers in Bioorganic Chemistry and Molecular Biology*, ed. Ovchinnikov, Yu. (Pergamon Press, Oxford), pp. 99–103.

15. Hruby, V. J. & Gehrig, C. A. (1986) *Medicinal Research Reviews* **9**, 343–401.
16. Hruby, V. J. (1974) in *Chemistry and Biochemistry of Amino Acids, Peptides and Protein*, ed. Weinstein, B., (Marcel Dekker, New York), Vol. 3, pp. 1–188.
17. Venkatachalam, C. M. (1968) *Biopolymers* **6**, 1425–1436.
18. Venkatachalam, C. M. & Ramachandran, G. N. (1969) *Annu. Rev. Biochem.* **38**, 45–82.
19. Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985) *Adv. Protein Chem.* **37**, 1–109.
20. Lewis, P. N., Momany, F. A. & Scheraga, H. A. (1973) *Biochim. Biophys. Acta* **303**, 211–229.
21. Ovchinnikov, Y. A. & Ivanov, V. T. (1975) *Tetrahedron* **31**, 2177–2209.
22. Kessler, H. (1982) *Angew. Chem., Int. Ed. Engl.* **21**, 512–523.
23. Urry, D. W. & Walter, R. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 956–958.
24. Brewster, A. I. R., Hruby, V. J., Glasel, J. A. & Tonelli, A. E. (1973) *Biochemistry* **12**, 5294–5304.
25. For a comprehensive review see Hruby, V. J. & Lebl, M. (1987) in *Handbook of Neurohypophyseal Analogs*, eds. Jost, K., Lebl, M. & Brtnik, F. (CRC Press, Boca Raton), Vol. I, part 1, pp. 105–155 and subsequent sections.
26. Veber, D. F. (1981) in *Peptides: Synthesis–Structure–Function*, eds. Rich, D. H. & Gross, E. (Pierce Chem. Co., Rockford), pp. 685–694 and references therein.
27. Sawyer, T. K., Sanfilippo, P. J., Hruby, V. J., Engel, M. H., Heward, C. B., Burnett, J. B. & Hadley, M. E. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5754–5758.
28. Sawyer, T. K., Hruby, V. J., Darman, P. S. & Hadley, M. E. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1751–1755.
29. Freidinger, R. M. & Veber, D. F. (1984) in *Conformationally Directed Drug Design*, eds. Vida, J. A. & Gordon, M. (ACS Symposium Series 251, Washington, D. C.), pp. 169–187.
30. Hruby, V. J., Al-Obeidi, F. A. & Kazmierski, W. (1990) *Biochem. J.* **268**, 249–262.
31. Al-Obeidi, F. A., Hadley, M. E., Pettitt, B. M. & Hruby, V. J. (1989) *J. Am. Chem. Soc.* **111**, 3413–3416.
32. Al-Obeidi, F. A., Castrucci, A. M. L., Hadley, M. E. & Hruby, V. J. (1989) *J. Med. Chem.* **32**, 2555–2561.
33. London, R. E., Stewart, J. M., Cann, J. R. & Matwiyoff, N. A. (1978) *Biochemistry* **17**, 2270–2277.
34. Nikiforovich, G. V., Rozenblit, S. A., Chipens, G. I. (1982) in *Chemistry of Peptides and Proteins*, eds. Voelter, W., Wunsch, E. Ovchinnikov, Y. & Ivanov, V. (Walter de Gruyter & Co., Berlin) Vol. 1, pp. 407–414.
35. Lenkinski, R. E., Stephens, R. L. & Krishna, N. R. (1981) *Biochim. Biophys. Acta* **667**, 157–165.
36. Sekacis, I., Shenderovich, M., Nikiforovich, G., Liepins, E., Polevaya, L. & Chipens, G. (1988) *Collect. Czech. Chem. Commun.* **53**, 2810–2824.
37. a) DiMaio, J. & Schiller, P. W. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7162–7166; b) Schiller, P. W. & DiMaio, J. (1982) *Nature (London)* **297**, 74–76.
38. DiMaio, J., Nguyen, T. M.-D., Lemieux, C. & Schiller, P. W. (1982) *J. Med. Chem.* **25**, 1432–1438.
39. Schiller, P. W., Nguyen, T. M.-D., Lemieux, C. & Maziak, L. A. (1985) *J. Med. Chem.* **28**, 1766–1771.
40. Mosberg, H. I., Hurst, R., Hruby, V. J., Galligan, J. J., Burks, T. F., Gee, K. & Yamamura, H. I. (1982) *Biochem. Biophys. Res. Commun.* **106**, 506–512.
41. Mosberg, H. I., Hurst, R., Hruby, V. J., Gee, K., Yamamura, H. I., Galligan, J. J. & Burks, T. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5871–5874.
42. Hruby, V. J., Kao, L. F., Pettitt, B. M. & Karplus, M. (1988) *J. Am. Chem. Soc.* **110**, 3351–3359.
43. Mosberg, H. I., Sobczyk-Kojiro, K., Subramanian, P., Crippen, G. M., Ramalingam, K. & Woodard, R. W. (1990) *J. Am. Chem. Soc.* **112**, 822–829.
44. Spatola, A. F. (1983) in *Chemistry and Biochemistry of Amino Acids, Peptides and Proteins*, ed. Weinstein, B. (Marcel Dekker, New York), Vol. 7, pp. 263–357.
45. Goodman, M. & Chorev, M. (1979) *Acc. Chem. Res.* **12**, 1–8.
46. Berman, J. M., Goodman, M., Nguyen, T. M.-D. & Schiller, P. W. (1983) *Biochem. Biophys. Res. Commun.* **115**, 864–870.
47. Kemp, D. S. & Stites, W. F. (1988) *Tetrahedron Lett.* 5057–5060.

48. Kemp, D. S. & Carter, J. S. (1987) *Tetrahedron Lett.* 4645–4648.
49. Kahn, M., Wilke, S., Chen, B. & Fujita, K. (1988) *J. Am. Chem. Soc.* **110**, 1638–1639.
50. Kahn, M. & Wilke, S. (1990) in *Peptides: Chemistry, Structure and Biology*, eds. Rivier, J. E. & Marshall, G. R. (ESCOM Publisher, Leiden), pp. 498–500.
51. Kemp, D. S., Boyd, J. G., Curran, T. P. & Fotouki, N. (1990), in *Peptides: Chemistry, Structure and Biology*, eds. Rivier, J. E. & Marshall, G. R. (ESCOM Publisher, Leiden), pp. 861–864.
52. Nicolaou, K. C., Salvino, J. M., Raynor, K., Pietramico, S., Reisine, T., Freidinger, R. & Hirschmann, R. (1990), in *Peptides: Chemistry, Structure and Biology*, eds. Rivier, J. E. & Marshall, G. R. (ESCOM Publisher, Leiden), pp. 881–884.
53. G. Olson, G. L., Voss, M. E., Hill, D. E., Kahn, M., Madison, V. S. & Cook, C. M. (1990) *J. Am. Chem. Soc.* **112**, 323–333.
54. Olson, G. L., Cheung, H.-C., Voss, M. E., Hill, D. E., Kahn, M., Madison, V. S., Cook, C. M., Spinwall, J. & Vincent, G. (1990) in *Proceeding, BIOTECH USA 1989* (Conference Management Corp., Norwalk), pp. 348–360.
55. Pelton, J. T., Gulya, K., Hruby, V. J., Duckles, S. P. & Yamamura, H. I. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 236–239.
56. Pelton, J. T., Kazmierski, W., Gulya, K., Yamamura, H. I. & Hruby, V. J. (1986) *J. Med. Chem.* **29**, 2370–2375.
57. Kazmierski, W., Wire, W. S., Lui, G. K., Knapp, R. J., Shook, J. E., Burks, T. F., Yamamura, H. I. & Hruby, V. J. (1988) *J. Med. Chem.* **31**, 2170–2177.
58. Kazmierski, W. & Hruby, V. J. (1988) *Tetrahedron* **44**, 697–710.
59. Kazmierski, W., Yamamura, H. I. & Hruby, V. J. (1991) *J. Am. Chem. Soc.* in press.
60. Hruby, V. J., Toth, G., Prakash, O., Davis, P. & Burks, T. F. (1989) in *Peptides 1988*, eds. Jung, G. & Bayer, E. (W. de Gruyter, Berlin), pp. 616–618.
61. Nikiforovich, G., Prakash, O., Gehrig, C. & Hruby, V. J., in preparation.

The Design of Novel Proteins

Bernd Gutte and Stephan Klauser

Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057
Zürich, Switzerland

Protein design comprises the modification of structure and function of naturally occurring proteins and the construction of artificial proteins. Although more difficult, the construction of artificial proteins or *de novo* design is gaining ground as shown by the most recent review of the field (1). Two approaches can be distinguished,

- (i) Attempts to model structural frameworks composed of α -helices, β -sheets, or both for subsequent introduction of a binding or catalytic function (2–9), and
- (ii) Attempts to design functional polypeptides in one step (10–17).

Design and structure elucidation of artificial proteins could become of considerable industrial, environmental, and medical interest and are also expected to make a strong contribution to our understanding of the mechanisms of protein folding. It is very encouraging that even in the absence of detailed knowledge of the protein folding rules a DDT-binding and degrading polypeptide (13, 18), an opioid receptor (15), and ion channels (16) could be designed.

In this article, we summarize the contributions from our laboratory to the field of protein design.

Materials and Methods

Polypeptide and Polydeoxyribonucleotide Synthesis and Purification

All polypeptides were synthesized by the solid phase method (19, 20) using *t*-Boc chemistry. They were deprotected and cleaved from the solid support by liquid HF at 0 °C during 1 h, extracted with 1 M acetic acid, lyophilized, dissolved in 0.05 M NH_4HCO_3 (if necessary in the presence of urea), stirred for 5 h, lyophilized again, and then purified by gel filtration, ion exchange, affinity and high performance liquid chromatography (HPLC).

Polydeoxyribonucleotides were synthesized by the phosphite method (21) on aminopropyl glass as solid support using deoxynucleoside phosphoramidites as reactive monomers (22). They were purified by gel electrophoresis on 20% cross-linked polyacrylamide in the presence of 7 M urea.

Determination of Dissociation Constants

Dissociation constants (K_D) of polypeptide–ligand complexes were determined by gel filtration (23), UV difference spectroscopy, CD (circular dichroism) spectral

titration with ligand in the near UV, and – in the case of the DNA-binding polypeptides – band shift experiments on gels.

CD Spectra

CD spectra were recorded on a Jasco J-500C spectropolarimeter. Peptide concentrations were determined by quantitative amino acid analysis after acid hydrolysis.

Results

A Synthetic 63-Residue Analogue of RNase A (Ribonuclease A)

The successful synthesis of RNase A in 1969 (24–26) opened the way for structure–function studies on proteins. The first examples for large-scale truncations of an enzyme seem to be the synthetic 70- and 63-residue analogues of RNase A (27–29). Based on the X-ray structure of the 124-residue natural enzyme (30), several superficial regions were omitted, leaving mainly the hydrophobic core that houses the largest part of the active site. The properties of the 63-residue analogue (28) (figure 1) were particularly interesting: it was bound by an affinity matrix specific for the active site fold of RNase A, it had RNase A-like substrate specificity and – in equimolar concentrations – up to ~ 10% of the activity of the natural enzyme. It also cross-reacted with anti-RNase A antibodies (unpublished results). All but one of the residues lying outside the energy level contours in the Ramachandran diagram of RNase S (30, 31) were absent from the 63-residue analogue or were terminal residues of segments of it.

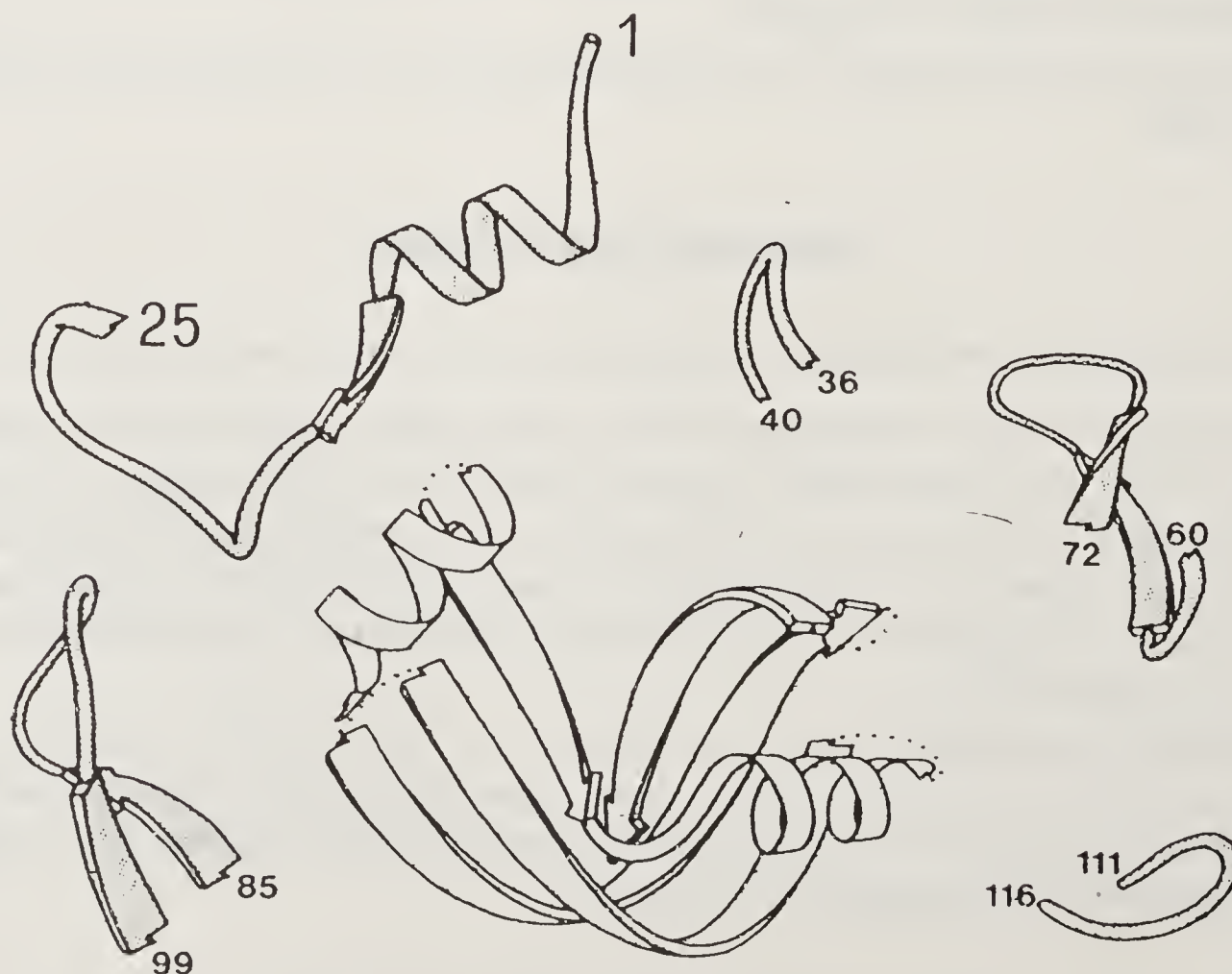


Figure 1. 63-residue RNase A analogue (bottom centre) surrounded by the segments that were omitted. Arrows are β -strands and spirals are α -helices. Reproduced with permission from ref. 1.

At the time, it was concluded that despite the drastic omissions of outer regions of the parent molecule the folding of the resulting 63-residue analogue was apparently very similar to that of the corresponding portions in the natural enzyme. Furthermore, all residues essential for substrate binding and catalysis seemed to be present in the truncated ribonuclease.

Taking advantage of today's advanced synthesis and purification procedures we have synthesized the 63-residue RNase A analogue again for CD, NMR, and crystallographic studies.

An Artificial 34-residue Polypeptide Shows Sequence-Specific Interaction with RNA

The successful preparation of an α -helical decapeptide with lysozyme-like activity by Dhar and coworkers (10) and a 63-residue polypeptide possessing RNase A-like activity (28) stimulated attempts to *design* a sequence-specific RNA-binding protein (11). The tools of design were secondary structure prediction (32,33) and model building. We used the trinucleotide GAA as ligand and assembled around it a polypeptide sequence (figure 2) that allowed the formation of a potential binding site. In this model, the major contribution to GAA binding was made by stacking interactions between the side chains of the aromatic residues in positions 1, 3 and 5 and the intercalating bases of the trinucleotide ligand.

After synthesis and purification, the artificial 34-residue polypeptide was characterized. We found that GAA was indeed bound but that the binding of CCC was ~ 100 times stronger (K_D , 5×10^{-6} M). The results of a CD spectral titration of the designed peptide with 2'-CMP (figure 3) confirmed the proposed stacking interactions between the aromatic amino acid residues and nucleotide bases. In control experiments, binding of aromatic amino acids by the 34-residue polypeptide was too weak to be detected. Unexpectedly, the designed peptide was enzymatically active; using tRNA as substrate a dimer of the peptide had 2.5% of the activity of RNase A. The substrate specificities of artificial and natural ribonuclease were similar but not identical.

Despite the discrepancy between expected and experimental ligand or substrate specificity the conclusion is allowed that functional polypeptides can be designed *de novo* using the approach described above (11).

Sequence-Specific DNA Binding of Designed Polypeptides – Toward Artificial Repressors

Replication and transcription are controlled by sequence-specific DNA-binding proteins (enzymes). Viral transcription, for example, is stimulated by enhancer-binding

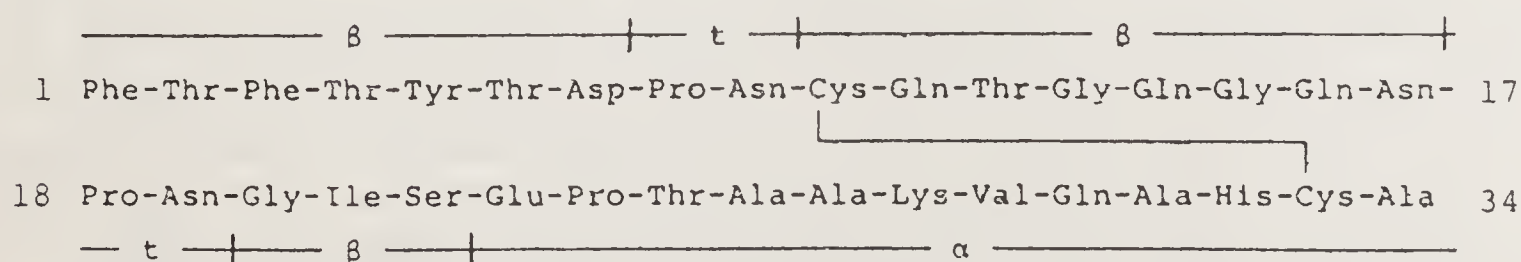


Figure 2. Sequence of the designed 34-residue polypeptide with weak ribonuclease activity. Bars indicate the extension of the proposed secondary structural elements (α , α -helix; β , β -strand; t, β -turn). A stabilizing disulfide bond was formed by half-cystines 10 and 33.

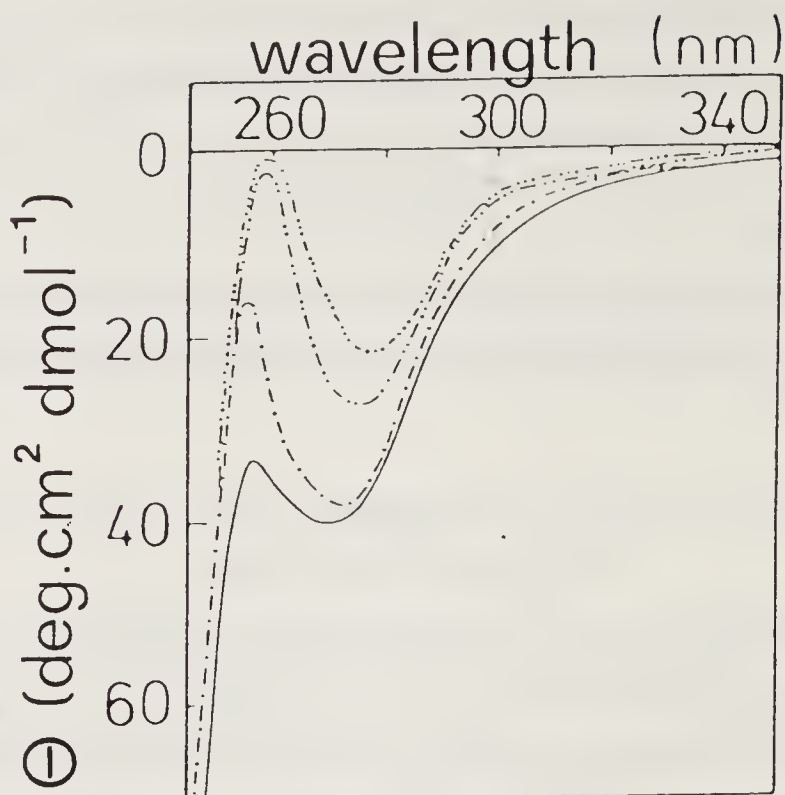


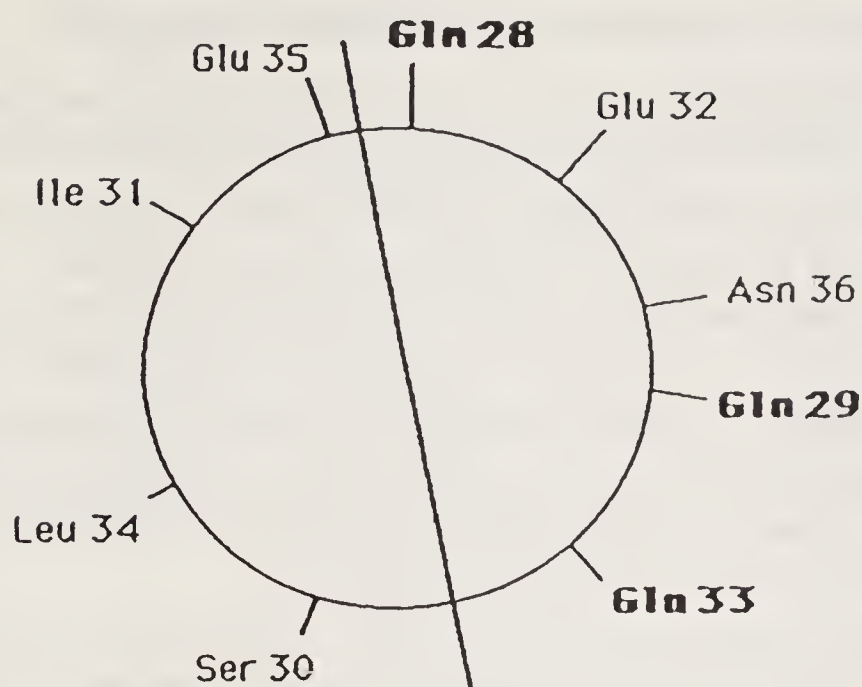
Figure 3. Near UV CD spectrum of dimeric 34-residue polypeptide (solid line; 9 μM peptide in 0.01 M sodium acetate, pH 5.3, at 22 °C) titrated with increasing concentrations of 2'-CMP (---, 6 μM ; - · - ·, 20 μM ; · · · ·, 60 μM). The spectra were corrected for the intrinsic ellipticity of the ligand.

proteins. This raises the possibility that polypeptides designed to interfere with the enhancer binding of these proteins block the transcription and replication of viruses.

To improve the chances of the design of a sequence-specific DNA-binding protein, we based our work on the X-ray structure of the repressor–operator complex of bacteriophage 434 (34, 35). In this complex, the helix–turn–helix motifs of the dimeric repressor recognize the dimeric symmetrical operator sequences ACAAG and ACAAT. The strongest contribution to the specificity of the interaction seems to come from the helical segment 28 to 36 (“recognition helix”) of 434 repressor which binds in the major groove of 434 operator. Glutamine residues 28, 29 and 33 that are proposed to make specific hydrogen bonds with operator bases (34, 35) are located on the DNA-binding side of the recognition helix (figure 4).

The enhancer region of HIV (human immunodeficiency virus) contains ACAAG and ACTTT sequences (36) which are identical or similar to the operator sequences of bacteriophage 434. However, in physiological conditions we could not detect binding between the helix–turn–helix motif of 434 repressor and synthetic oligodeoxyribonucleotides containing monomeric ACAAG, ACTTT, or both. Binding was observed (K_D , 10^{-8} to 10^{-9} M) when eight positively charged amino acid residues were added to each end of the recognition helix. Despite the many new positive charges, the specificity of the interaction remained almost unchanged (37). The exchange of residues of the recognition helix (for example, Glu 32 Arg) (figure 4), however, usually had a negative effect on the binding specificity. Figure 5 showed that the affinity of the Arg 32 mutant for calf thymus DNA was much higher than that of the polypeptide with unaltered sequence of the recognition helix (38).

NMR studies of the solution structure of our polypeptide–DNA complexes are under way and may provide important information for redesigning our artificial sequence-specific DNA-binding proteins. In the meantime various ways of introducing these proteins into healthy and HIV-infected cells will be investigated.



DNA-binding surface

Figure 4. Helical wheel representation of the recognition helix (residues 28 to 36) of bacteriophage 434 repressor. The DNA-binding side is on the right.

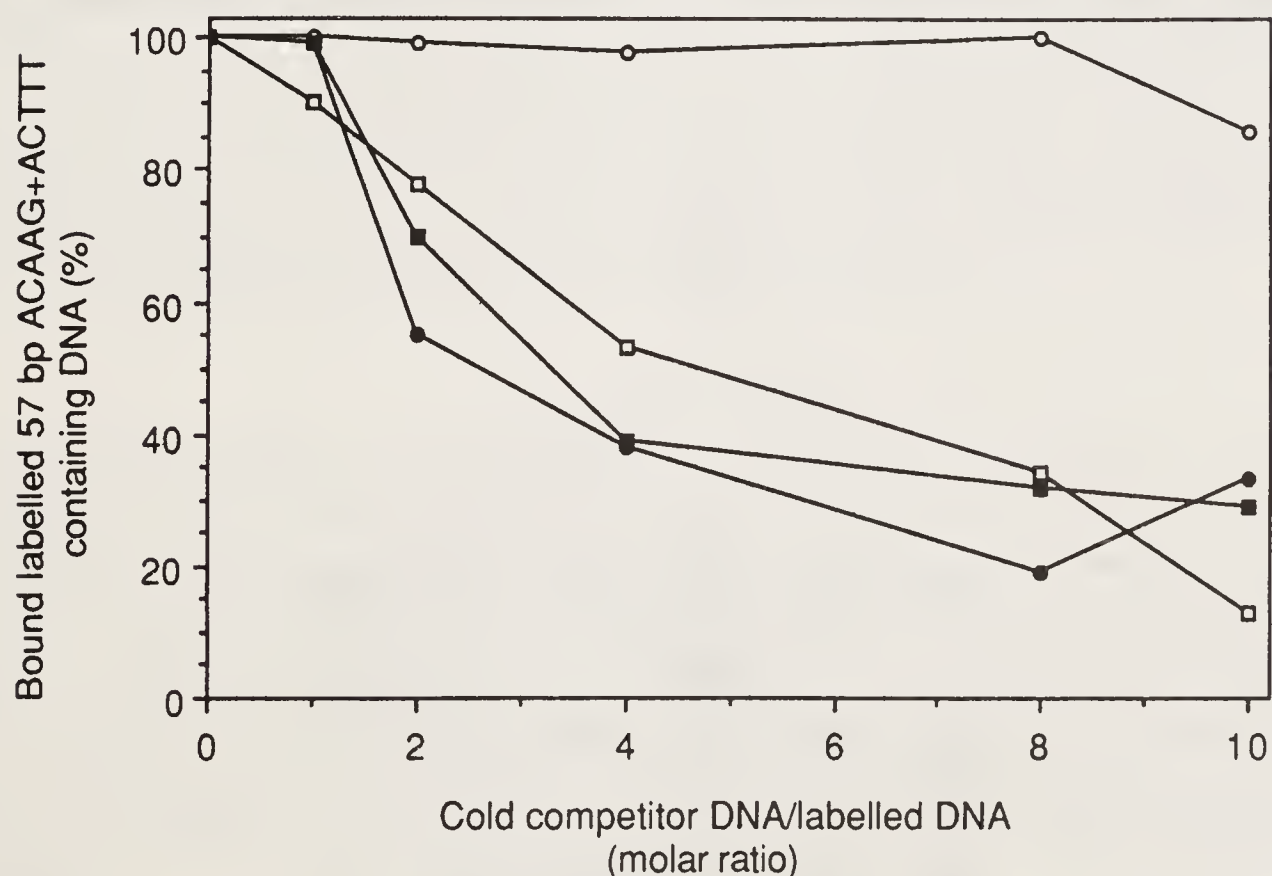


Figure 5. Gel retention assays of the specificity of DNA binding. Incubation of 10^{-8} M 5'-labelled 57-bp ACAAG + ACTTT-containing DNA with 10^{-6} M polypeptide was performed for 15 min at 37°C in 100 mM $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$, pH 7.5, 60 mM NaCl, 15 mM KCl, 4 mM MgCl_2 . Then competitor DNA was added to the preformed DNA-polypeptide complexes and incubation was continued for another 15 min at 37°C. Samples were analyzed on 8% cross-linked polyacrylamide gels in 22.5 mM Trisborate, pH 7.5. The amount of complex containing labelled DNA was determined by autoradiography. Competitor DNAs were cold 57-bp ACAAG + ACTTT-containing DNA (□, Glu 32 polypeptide; ■, Arg 32 mutant) and cold sonicated calf thymus DNA (○, Glu 32 polypeptide; ●, Arg 32 mutant).

Design of a 24-Residue DDT-Binding Polypeptide

The design of an artificial DDT-binding polypeptide seemed to be interesting *per se*. If, in addition, a natural DDT-binding protein could be isolated, the elucidation and comparison of the structure of the ligand binding site of the two proteins would greatly aid all future work in the field of protein design in general.

The construction of the 24-residue DDT-binding polypeptide (13) followed the approach described in ref. 11. Figure 6 shows the proposed structure of the polypeptide-DDT complex. CD studies and DDT-binding properties of the 24-residue polypeptide and of analogues indicated a high degree of β -sheet content and lent support to the proposed mode of ligand binding (39). Quantitative structural data are expected to be obtained from 2D NMR measurements.

The designed 24-residue polypeptide bound DDT ~ 1000 times more strongly than serum albumin and ~ 70 times more strongly than β -casein, two natural DDT-binding proteins. The dissociation constant of the 24-residue peptide-DDT complex was 0.8×10^{-6} M (18). On the other hand, there was very little if any binding of aromatic ligands such as dioxin, heme, or nucleotides (39).

Naturally, we were not only interested in DDT binding but also in DDT degradation. We found that the cytochrome P-450 model system *hemin + excess cysteine* (40) was able to degrade the insecticide DDT partially and that the rate of degradation was enhanced in the presence of designed 24-residue polypeptide and β -casein (the catalyzed DDT metabolism was $\sim 3 \times 10^5$ times faster than the uncatalyzed) (18). DDT degradation is visualized in figure 7. The major products

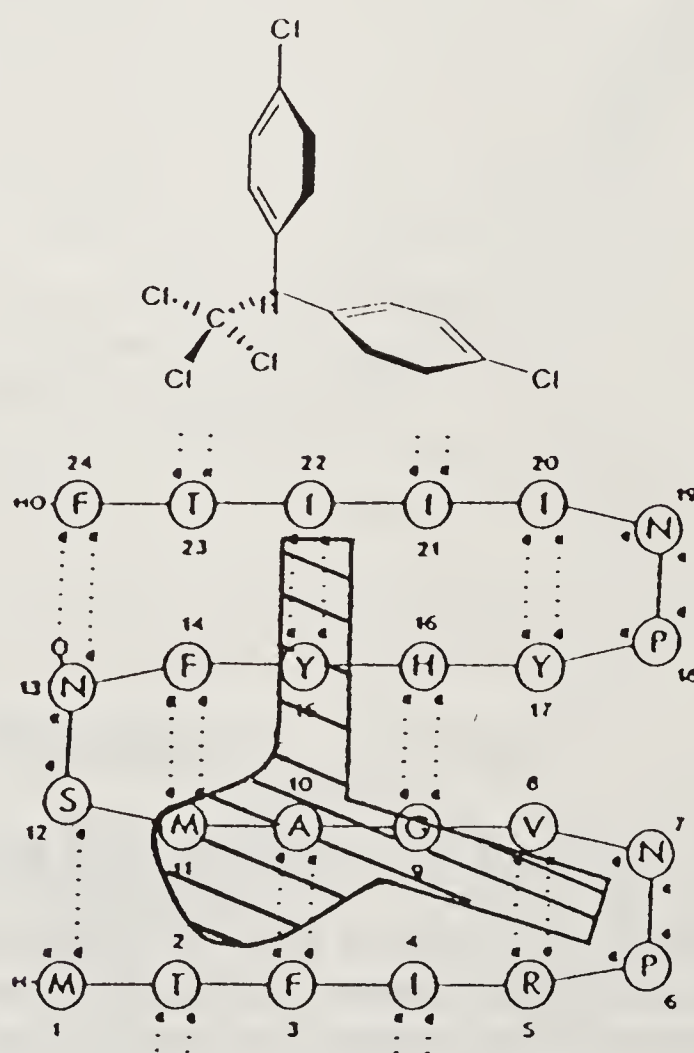


Figure 6. Amino acid sequence and proposed folding of a designed 24-residue DDT-binding polypeptide. The structural formula of DDT is shown on top. The position of a potential DDT binding site which, in the model, is highly complementary to the ligand is indicated by the hatched area. Dotted lines are hydrogen bonds of the β -pleated sheet.

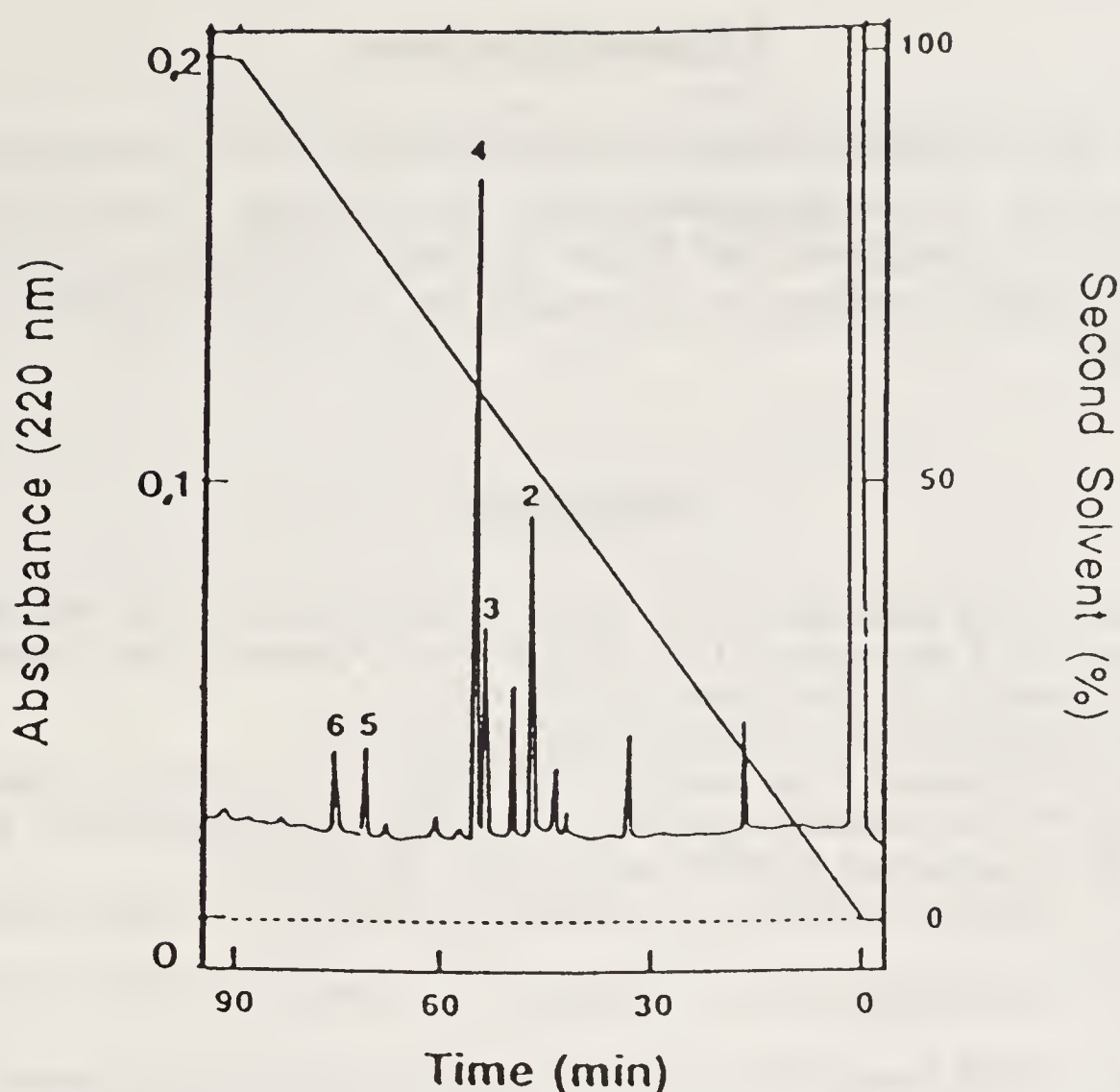


Figure 7. Resolution of the products of hemin–cysteine–polypeptide-catalyzed DDT degradation by reversed-phase HPLC. First solvent, 10% CH_3CN in 0.1% aqueous TFA (trifluoroacetic acid); second solvent, 75% CH_3CN in 0.1% aqueous TFA. The reaction conditions were as follows: 2.5 mM DDT, 0.34 mM hemin, 68 mM cysteine and 0.1 mM designed 24-residue DDT-binding polypeptide in 0.05 M NH_4HCO_3 , pH 7.7/ethanol (5:6, v/v), 6 h, 37°C. Peaks 2–4 are DDT metabolite–cysteine conjugates, peak 6 is unreacted DDT.

were three water-soluble, non-toxic conjugates of DDT metabolites with cysteine which had lost two or three of the five chlorine atoms of DDT per molecule and whose structures were elucidated by gas chromatography/mass spectrometry (18). It seemed that the hemin–cysteine–polypeptide mixtures or complexes were primitive enzymes in which the polypeptides (the designed 24-residue peptide and β -casein) served to improve the solubility of DDT for the hemin–cysteine-catalysed reaction.

A β -galactosidase–DDT-binding-peptide fusion protein, produced in *Escherichia coli*, also bound DDT and accelerated its degradation (14). A β -galactosidase control was negative.

Finally, we have prepared monoclonal anti-DDT antibodies (41) to compare structure and properties of designed and natural DDT-binding proteins.

Conclusion

De novo design of proteins is a fascinating new field whose limits are mainly set by the courage of the researcher and the willingness of funding agencies to support the work.

Acknowledgements

We thank Mrs. I. Willimann for excellent secretarial work. The financial support by the Deutsche Forschungsgemeinschaft, the Stiftung Volkswagenwerk, the Schweizerische Nationalfonds, the Roche Research Foundation, the Stiftung zur Förderung der AIDS-Forschung in der Schweiz, and the Kanton of Zürich is gratefully acknowledged.

References

1. Richardson, J. S. & Richardson, D. C. (1989) *Trends Biochem. Sci.* **14**, 304–309.
2. Richardson, J. S. & Richardson, D. C. (1987) in *Protein Engineering*, eds. Oxender, D. L. & Fox, C. F. (Alan R. Liss, New York), pp. 149–163.
3. Regan, L. & DeGrado, W. F. (1988) *Science* **241**, 976–978.
4. Erickson, B. W., Daniels, S. B., Reddy, P. A. & Higgins, M. L. (1988) in *Advances in Gene Technology, Protein Engineering and Production*, ICSV, Short Reports, Vol. 8, pp. 4–5.
5. Mutter, M. & Vuilleumier, S. (1989) *Angew. Chem.* **101**, 551–571.
6. Karle, I. L., Flippen-Anderson, J. L., Uma, K. & Balaram, P. (1989) *Biochemistry* **28**, 6696–6701.
7. Hecht, M. H., Richardson, D. C., Richardson, J. S. & Ogden, R. (1989) *J. Cell. Biochem.* **13A**, 86.
8. Hubbard, T. J. P. & Blundell, T. L. (1989) in *Computer Simulations of Biomolecular Systems: Theoretical and Experimental Applications*, eds. van Gunsteren, W. F. & Weiner, P. K. (ESCOM, Leiden), pp. 168–182.
9. Goraj, K., Renard, A. & Martial, J. A. (1990) *Protein Eng.* **3**, 259–266.
10. Chakravarty, P. K., Mathur, K. B. & Dhar, M. M. (1973) *Experientia* **29**, 786–788.
11. Gutte, B., Däumigen, M. & Wittschieber, E. (1979) *Nature (London)* **281**, 650–655.
12. Fukushima, D., Kupferberg, J. P., Yokoyama, S., Kroon, D. J., Kaiser, E. T. & Kézdy, F. J. (1979) *J. Am. Chem. Soc.* **101**, 3703–3704.
13. Moser, R., Thomas, R. M. & Gutte, B. (1983) *FEBS Lett.* **157**, 247–251.
14. Moser, R., Frey, S., Mürger, K., Hehlhans, T., Klauser, S., Langen, H., Winnacker, E.-L., Mertz, R. & Gutte, B. (1987) *Protein Eng.* **1**, 339–343.
15. Kullmann, W. (1984) *J. Med. Chem.* **27**, 106–115.
16. Lear, J. D., Wasserman, Z. R. & DeGrado, W. F. (1988) *Science* **240**, 1177–1181.
17. Sasaki, T. & Kaiser, E. T. (1989) *J. Am. Chem. Soc.* **111**, 380–381.
18. Langen, H., Epprecht, T., Linden, M., Hehlhans, T., Gutte, B. & Buser, H.-R. (1989) *Eur. J. Biochem.* **182**, 727–735.
19. Merrifield, R. B. (1963) *J. Am. Chem. Soc.* **85**, 2149–2154.
20. Barany, G. & Merrifield, R. B. (1980) in *The Peptides*, eds. Gross, E. & Meienhofer, J., (Academic Press, New York), Vol. 2, pp. 1–284.
21. Letsinger, R. L. & Lunsford, W. B. (1976) *J. Am. Chem. Soc.* **98**, 3655–3661.
22. Beaucage, S. L. & Caruthers, M. H. (1981) *Tetrahedron Lett.* **22**, 1859–1862.
23. Hummel, J. P. & Dreyer, W. J. (1962) *Biochim. Biophys. Acta* **63**, 530–532.
24. Gutte, B. & Merrifield, R. B. (1969) *J. Am. Chem. Soc.* **91**, 501–502.
25. Hirschmann, R., Nutt, R. F., Veber, D. F., Vitali, R. A., Varga, S. L., Jacob, T. A., Holly, F. W. & Denkwalter, R. G. (1969) *J. Am. Chem. Soc.* **91**, 507–508.
26. Gutte, B. & Merrifield, R. B. (1971) *J. Biol. Chem.* **246**, 1922–1941.
27. Gutte, B. (1975) *J. Biol. Chem.* **250**, 889–904.
28. Gutte, B. (1977) *J. Biol. Chem.* **252**, 663–670.
29. Gutte, B. (1978) *J. Biol. Chem.* **253**, 3837–3842.
30. Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B. & Richards, F. M. (1970) *J. Biol. Chem.* **245**, 305–328.
31. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283–437.
32. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47**, 45–148.

33. Levitt, M. (1978) *Biochemistry* **17**, 4277–4285.
34. Anderson, J. E., Ptashne, M. & Harrison, S. C. (1987) *Nature (London)* **326**, 846–852.
35. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988) *Science* **242**, 899–907.
36. Franza, B. R., Jr., Josephs, S. F., Gilman, M. Z., Ryan, W. & Clarkson, B. (1987) *Nature (London)* **330**, 391–395.
37. Hehlhans, T., Stolz, M., Cui, T., Mohajeri, H., Salgam, P., Leiser, A., Klauser, S., Mertz, R. & Gutte, B. (1990) *Experientia* **46**, A3.
38. Hehlhans, T., Stolz, M., Mohajeri, H., Salgam, P., Cui, T., Leiser, A., Städler, K., Siskawich, M., Mertz, R., Klauser, S. & Gutte, B., in preparation.
39. Klauser, S., Gantner-Holenstein, D., Salgam, P., Langen, H. & Gutte, B., in press.
40. Sakurai, H. (1980) *Chem. Pharm. Bull.* **28**, 3437–3439.
41. Bürgisser, D., Frey, S., Gutte, B. & Klauser, S. (1990) *Biochem. Biophys. Res. Commun.* **166**, 1228–1236.

Template Assembled Synthetic Proteins (TASP): A Chemical Approach to the Construction of New Proteins

Robert I. Carey and Manfred Mutter

Séction de Chimie de l'Université de Lausanne, Rue de la Barre 2,
CH-1005 Lausanne, Switzerland

Introduction

Peptide and protein molecules occur ubiquitously in nature where they mediate countless processes of recognition, binding, and catalysis. Chemists and biologists alike have long been fascinated by the many functions which globular proteins are capable of displaying, and consequently, many innovative attempts to mimic these properties have been made. Previous endeavors have included the preparation of polymeric catalysts as enzyme models (1) and the preparation and study of carefully designed rigid organic frameworks modeled after the catalytic sites of enzymes (2–4). In the past decade, however, our ability to study proteins has been remarkably increased due to the dramatic advances which have been made in the field of molecular biology (5, 6) and in the methodology that is commercially available for peptide synthesis and purification (7, 8). By means of recombinant DNA techniques it is now possible to synthesize almost any primary sequence consisting of the twenty naturally occurring amino acids, provided that the gene can be cloned and suitably expressed. Extension of this technique to the incorporation of unnatural amino acids is an area of current investigation (9). Total synthesis as a tool for studying small proteins has also been demonstrated in a very impressive fashion (10) in the preparation of dozens of analogues of human transforming growth factor α , TGF- α , a 50 amino acid residue protein which is very difficult to isolate from natural sources. Concurrent with these recent developments in the preparation of new proteins, the biophysical techniques available to study biological macromolecules have likewise rapidly advanced. Of special importance has been the increasingly widespread use of high field NMR spectroscopy (11). Also our ability to monitor and quantify small differences in protein structure (12) and activity (13, 14) is becoming increasingly more routine and more powerful. Taken together, the result of these advances has been a flood of new data on protein molecules, and it has made possible systematic studies of protein structure and function that could have only been hoped for a short time ago (15).

Protein *de novo* Design and the TASP Concept

With the knowledge of protein structure which we have been able to accumulate over the past few years and which we continue to accumulate at a very rapid pace,

it is now appropriate to ask if it is possible to design from first principles "*de novo*" and synthesize completely new protein molecules which contain amino acid sequences nonhomologous to those occurring in nature, but that exhibit all the characteristic properties of natural proteins, the most important of these properties being the ability to fold into a stable, predictable tertiary structure (16–22).

A descriptive account of the earliest attempts at *de novo* design of proteins (23–29) was the subject of a recent review (22), and the design principles on which these molecules were constructed have also been reviewed (16, 21, 30). While most of these early designs had circular dichroism (CD) and infrared (IR) data that were consistent with the proposed structure, in no case has definite structure proof been provided. Nevertheless, these pioneering studies established a sound foundation on which future studies can be based.

A common feature that the early designs shared was their linear mode of chain connectivity. Although assembling artificial proteins in a linear fashion (analogous to protein chain assembly in nature) is a logical synthetic direction to follow, any new protein designed in this way must still overcome the critical hurdle of folding into a well-defined, globular structure. However, despite our rapidly improving understanding of the processes involved in protein folding, we are still unable to predict with confidence tertiary structure from a primary sequence of amino acids. The severity of this problem in relation to *de novo* protein design is perhaps best illustrated by the vanishingly small number of primary sequences that are known to adopt a globular folded structure in nature compared to the almost infinite number of possibilities of folding which could exist. Therefore, until we have gained a better understanding of the complex folding mechanisms through which globular structures are formed by linear polypeptide chains, it is unlikely that even the most sophisticated model systems constructed in a linear manner will consistently contain all the information necessary to fold into the proposed tertiary structure. Successful designs are more likely to be single-hit phenomena rather than part of a successful general strategy. This difficult situation, often referred to as the protein folding problem is represented schematically in figure 1 and has been discussed previously in much detail (21).

These considerations have therefore led us to adopt an alternative approach to protein design (19), Template Assembled Synthetic Proteins (TASP), in which amphiphilic secondary structure blocks are attached to a conformationally directing template molecule. The anticipated effect of the template is to hold the attached blocks in close proximity, to enhance secondary structure formation, and to facilitate an intramolecular collapse of the amphiphilic peptide blocks to form a globular structure either before or after secondary structure formation has actually begun to form (for a discussion of the current hypotheses for the mechanism of protein folding, see 31, 32).

A number of the possible topologies for TASPs are shown in figure 2. Artificial proteins constructed within this branched-chain, unnatural architecture should have many advantages in comparison to their linear analogues in which the amphiphilic peptide blocks are connected through loop or turn sequences and which must reach their final folded state through a more stringently demanding pathway(s). Through the template effects described above, the TASPs should display an increased tendency to adopt a globular folded structure and a decreased tendency to succumb to *intermolecular* aggregation, a process which can fatally disrupt the *intramolecular* association that is required for proper folding. Furthermore, the accessibility of TASP

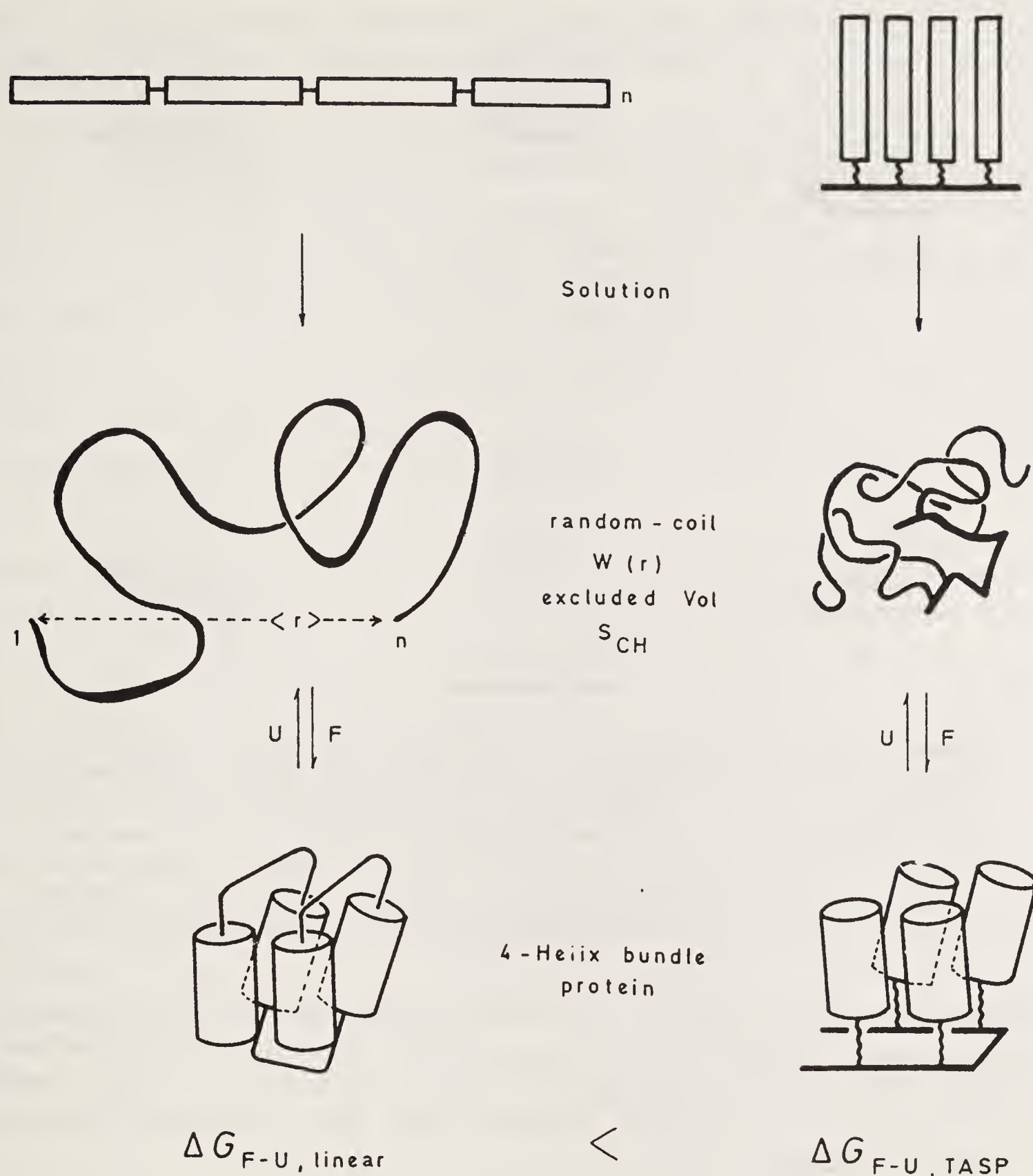


Figure 1. Schematic representation of thermodynamic aspects of the folding of linear (L) and branched (T) polypeptides. While the unordered (U) conformation of both molecules under poorly solvating conditions (near the so-called theta point) is exclusively determined by short range interactions, extensive “excluded volume effects” arise as a consequence of long range interactions, when the chain enjoys good solvation. This can be expressed in terms of an increase of the end-to-end vector $\langle r \rangle$. Due to the higher coil density in the branched molecule, the number of accessible conformations is smaller than in the linear molecule, i.e. the chain entropy, S_{chain} , is decreased. This in turn leads to a destabilization of the unfolded state T_u of the branched molecule. Under the assumption that the intramolecular interactions in the folded state (F) are similar in both linear and branched molecules (i.e., ΔH^* , $\Delta S^* = 0$), the branched molecule will be stabilized relative to the linear by a difference in the free energy $\Delta G_{L-T}^* = \Delta S_{\text{chain}}^*$ as can be taken from the figure. TASP molecules should therefore be more stable thermodynamically than linear polypeptides of equivalent packing arrangement (here, a four-helix bundle structure). This aspect of the energetics is of great importance with respect to subsequent modification of the structure in order, for example, to introduce functionality into the molecular structure. Taken from ref. 21.

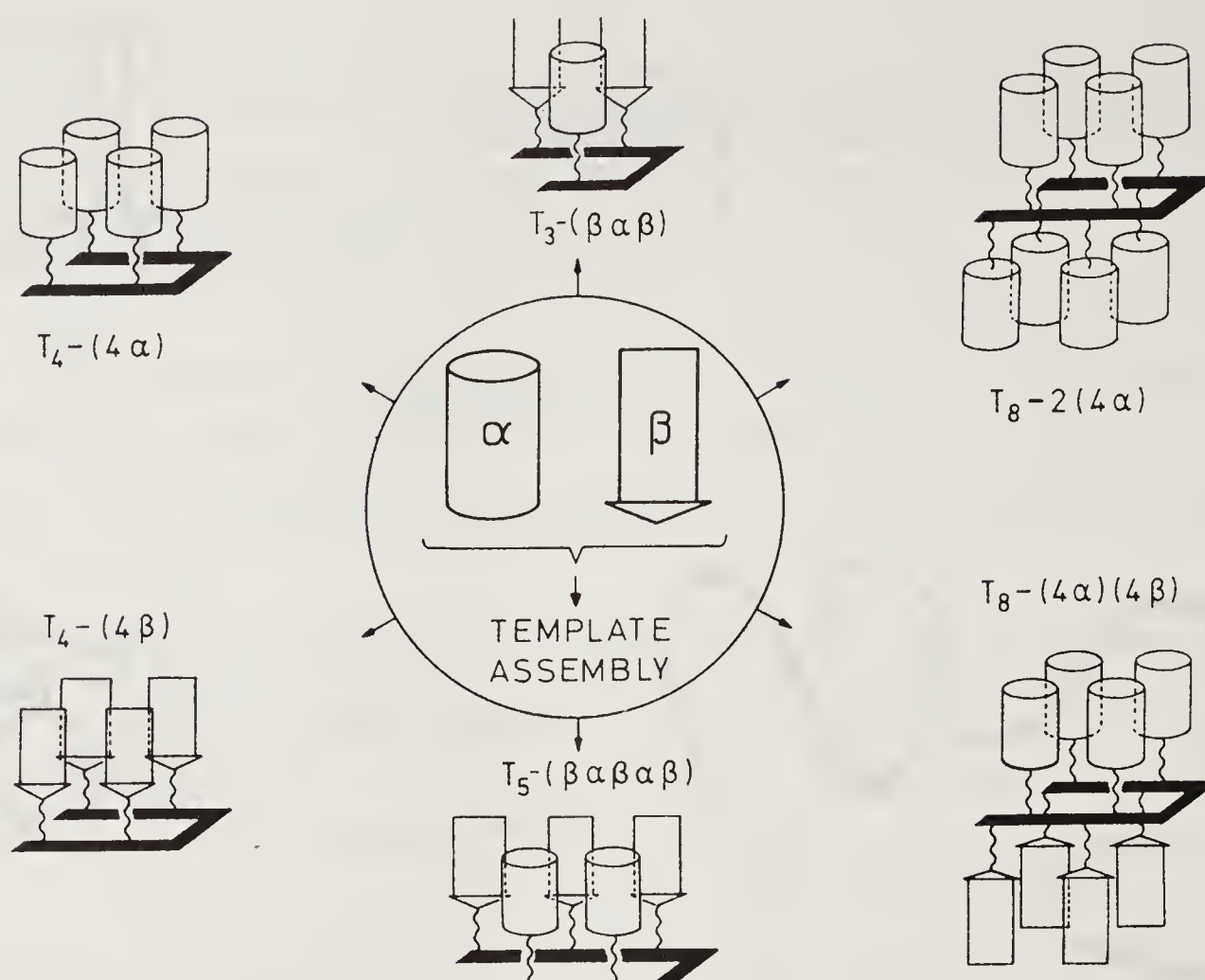


Figure 2. Schematic representation of different possible packing arrangements (topologies) in TASP molecules (T_n , T = template molecule, n = number of attachment sites on T ; α , β , amphiphilic peptide blocks with tendency for helix (α) and β -sheet (β) secondary structure formation). The number, nature, and spatial orientation of the attachment sites in the carrier molecule predetermine the topological features of the TASP molecule and thus the carrier acts as a template for the folding of the molecule into the desired tertiary structure. Taken from ref. 21.

molecules by chemical synthesis, the pertinent details of which will be discussed in the next section, brings forth the exciting possibility, to incorporate unnatural amino acids with widely varying steric and electronic properties into our target structures (e.g. the inclusion of amino acids bearing strongly acidic or basic side chains into hydrophobic binding pockets).

Construction of TASP Molecules—Synthetic Considerations

The chemical synthesis of protein molecules is a goal that has fascinated organic chemists for nearly a century. Although the accomplishments in this endeavour have been monumental (33), the point has hardly been reached at which one can say that the synthesis of a protein containing 100 or more amino acids is a facile or consistently routine task. To the synthetic chemist, TASP molecules represent a considerable challenge which increases with the size and complexity of the desired TASP structure. The nonlinear chain architecture, which is the feature that makes TASP molecules attractive from the viewpoint of protein folding, also makes TASP molecules inaccessible to synthesis by means of recombinant DNA (biological) techniques. The approach that has been taken initially by our laboratory has been that of stepwise solid phase synthesis using orthogonal protecting schemes (8, 33). Starting from templates of the general formula P-Lys-X-Lys-Pro-Gly-Lys-Y-Lys-Z (P, X, Y, Z = Ac

Lys, Glu, NH_2 , Ac = acetyl), TASPs with $\beta\alpha\beta$, 4- α , and 4- $\alpha/4\text{-}\beta$ topology have been synthesized by stepwise solid-phase methods and their conformational properties evaluated (21, 34–36).

A further set of TASPs, all of which bear a cyclic (via S–S bond formation) template and 4- α topology, has been synthesized more recently (37), and the general structural properties of these will be discussed in the following section. Whereas the solid phase synthesis has been routine, the purification and characterization of our products have been the subject of careful study (38). As expected, we have found that the larger and more complex the TASP, the more it becomes necessary to employ orthogonal techniques of chromatography to remove the cumulative impurities that result from solid phase synthesis and penultimate acidolytic deblocking. For the TASPs shown in figures 3 and 6, we have found that the application of preparative ion exchange chromatography after preparative HPLC has been very effective in removing impurities that would have otherwise remained undetected.

The strategy of TASP synthesis that involves solution phase condensation of protected peptide fragments with a given template molecule is also being pursued in our laboratory and to date has resulted in the successful preparation of 4- α TASP (39) in which the template molecule was a cyclic peptide containing two β -turn mimetics (40). While this strategy has been greatly facilitated by the developments both in classical solution phase techniques (41) and in the commercial availability of solid phase resins (42,43) that allow for the cleavage of fully protected peptide fragments, one must be careful to recognize the inherent limitations of amide bond forming processes in which the affinity for attachment involves a weakly basic peptide amine and an electrophilic carbon of an activated acyl derivative (see ref. 44 for a thorough discussion of this topic). Certainly entropy and dilution will set a limit to the size of molecules which can be coupled efficiently in any bimolecular process, and the necessity in TASP synthesis to couple usually at least four peptide blocks to a single template, further intensifies the problem.

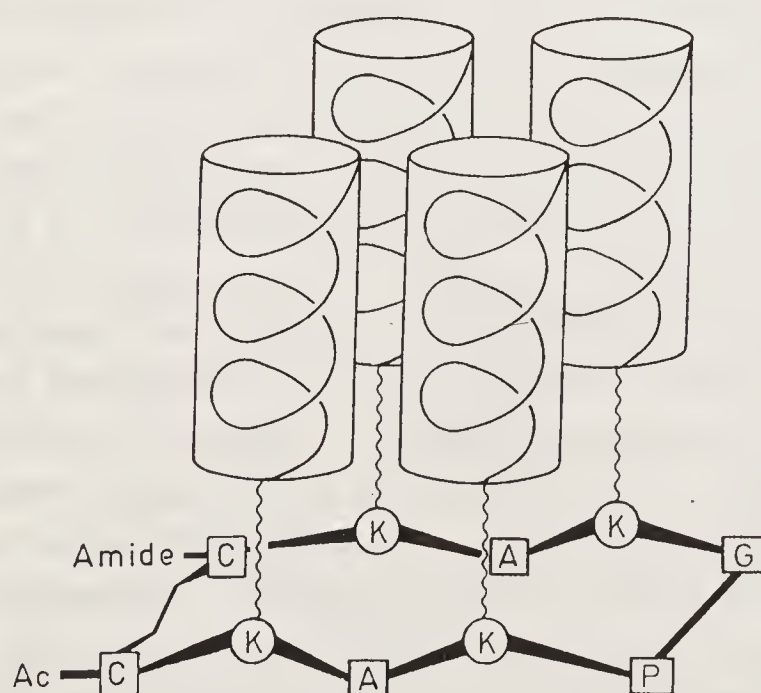


Figure 3. Schematic representation of the proposed 4 helix bundle structure for a recently synthesized TASP molecule (37) containing the template Ac-Cys-Lys-Ala-Lys-Pro-Gly-Lys-Ala-Lys-Cys- NH_2 which has been cyclized via disulfide (S–S) bond formation between the two cysteine residues. The four helical blocks are identical and have the sequence X-Glu-Ala-Leu-Glu-Lys-Ala-Leu-Lys-Glu-Ala-Leu-Ala-Lys-Leu-Gly (X = H, TASP; X = Ac, TASP'; see text).

Alternative approaches to the coupling of peptide blocks with template molecules or to the ligating of two or more TASP's to form multidomain artificial proteins are desirable and will probably be necessary for the efficient, reproducible synthesis of larger (> 100 amino acids) target designs. As state-of-the-art enzymatic synthesis (45, 46) does not appear to be sufficiently powerful to meet our present requirements, attempts are underway in our laboratory to utilize the unconventional method of prior thiol capture (47, 48), which is especially attractive in that it allows the coupling of peptides which are either fully or minimally protected. Recently, this methodology has been successfully employed for the preparation of two peptides (39 and 25 amino acids long, respectively) through the coupling ($13 + 26$ and $13 + 12$) of peptide fragments in which all of the amino acid side chain protecting groups had been removed with the exception of those on the thiols of the cysteine residues which were appropriately functionalized for their role in prior thiol capture (49). Successful application of this methodology to TASP design will offer the further advantageous feature that the coupling step generates a nascent cysteine thiol to which spectroscopic probes, enzymatic cofactors, or various steric and functional groups may be attached. Still other methods for the attachment of peptides to a template are presently under development. For example, after proper functionalization of the template and peptide block, one can envision efficient ligation processes proceeding through thioether formation or a Diels–Alder reaction. Such nonpeptidic linkages may prove to be advantageous both from a structural and synthetic standpoint.

Properties of TASP Molecules

By design, template-assembled synthetic proteins are macromolecules which exhibit structural features of both synthetic polymers (branched chains, grafted polymers) and globular proteins (peptide building blocks, hydrophobic core, hydrophilic surface) (19), and it is significant to note that all of the TASP molecules described to date have been soluble in aqueous buffers, a property that had been predicted for TASP's and that suggests the formation of a globular structure containing a hydrophobic core and a hydrophilic surface. Conformational analysis of the TASP molecule (37) whose proposed structure is shown in figure 3 strongly suggests that this structure is indeed attained. The CD spectrum of this molecule ($X = H$, cf. figure 3) in aqueous buffer, pH 7, is shown in figure 4 together with the single helical block (attached to the ϵ -amino group of Ac-Lys-NH₂) for a comparative reference. Both the reference peptide block and the TASP molecule have an unblocked and therefore charged amino terminus. Greatly enhanced helicity is observed for the peptide block when it is attached to the template, even though this TASP design contains two destabilizing features for a 4-helix bundle structure, the antiparallel helical arrangement (50) and the positively charged N-termini of the helices (51). Interestingly, acetylation of the N-termini of the template bound helices does not result in any significant change of the CD spectrum depicted in figure 4, while the acetylation of the single block leads to a drastic increase in helicity. Guanidine hydrochloride (guaHCl) induced denaturation of the N-terminally acetylated TASP' ($X = Ac$, cf. figure 3) was followed as shown in figure 5 by the decrease in ellipticity at 222 nm, and the helical structure of the TASP' was found to be stable to concentrations of guaHCl up to 3–4 M, a stability comparable to that of natural proteins (52).

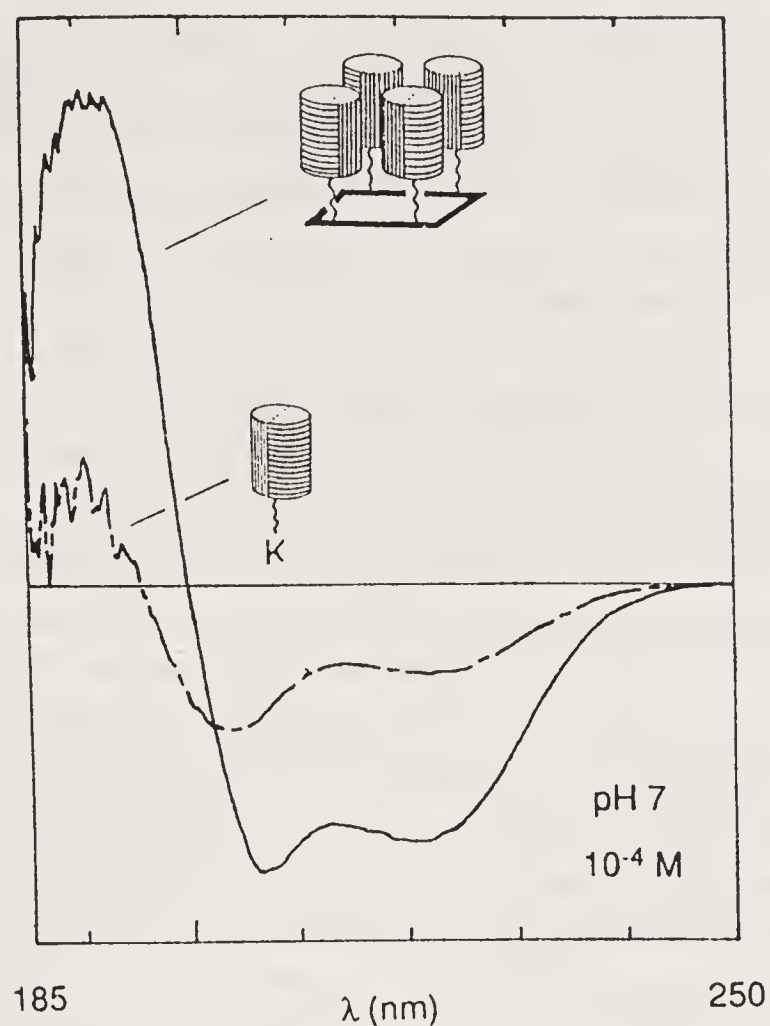


Figure 4. CD spectra in H_2O at pH 7 of TASP ($X = \text{H}$, cf. figure 3) (—) and the corresponding helix block (attached to the ϵ -amino group of Ac-Lys- NH_2) (----).

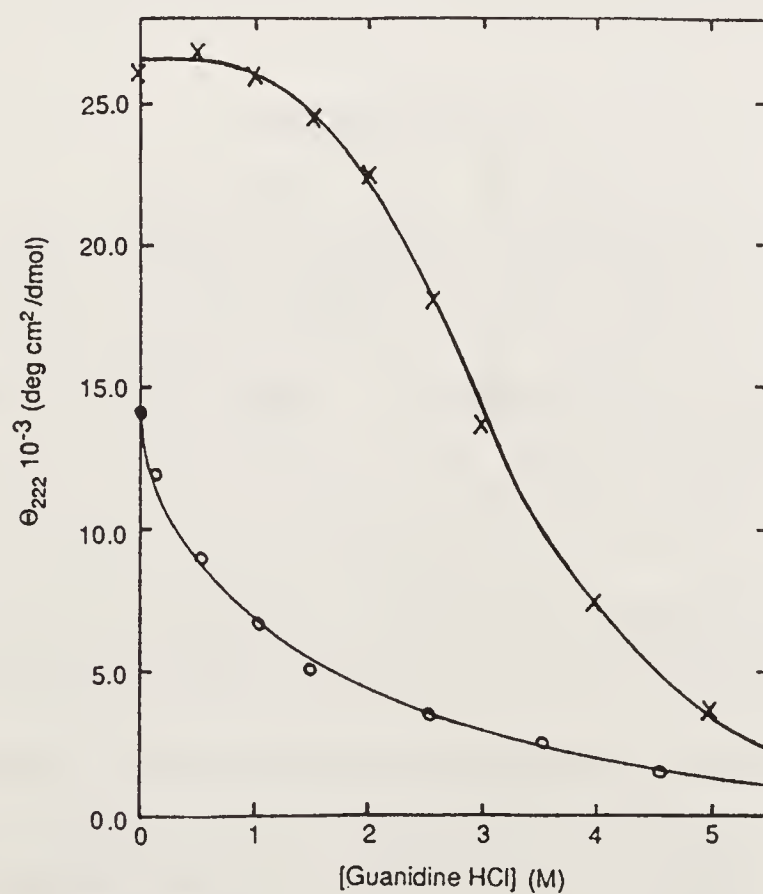


Figure 5. Denaturation of TASP' ($X = \text{Ac}$, cf. figure (8) (+)) by increasing concentrations of guanidine hydrochloride (guaHCl). The curve for the single helix block is shown for comparison (O).

Further structural data has been obtained through the synthesis of the 4α -helix bundle TASP depicted in figure 6, in which there are two different helices attached to the cyclic template (53). Analogous to the TASPs studied before it, this TASP with two separate helices possessed 1D NMR spectra and CD spectra that were fully consistent with the predicted fold and that indicated strongly that the attachment of the peptide blocks to the template changes their conformation. Although no attempt is being made here to characterize this folding in any detail (54), a preliminary and qualitative assessment of the compact conformation of $T_4-(3\alpha';\alpha)$ is however available from the three leucine δ or δ' to Phe 8 cross peaks shown in figure 7. At least two leucines must be involved, possibly more, as there is likely to be a chemical shift degeneracy between the δ or δ' methyls of the same leucines or between different leucines occupying the same sequence positions in the three α' chains of $T_4-(3\alpha';\alpha)$. All leucines within the primary sequence are distant from Phe8, and in the absence of a fold it would not be expected that the NOEs in figure 7 would be observed. However, given the presumed helical fold, interchain contacts between the Phe 8 of the α chain and the Leu 8 of the α' chain, and possibly also intrachain contacts between Leu 4 and Phe 8 of the α -chain would be expected.

Although we are far from having realized all of the many goals for the *de novo* design of proteins that we established for ourselves at the outset of our research in this area, the syntheses and conformational studies accomplished so far have strongly supported our belief that the TASP concept represents a very promising approach to the design of novel proteins. It should be mentioned here that our approach has been used in the preparation of a heme containing artificial protein (55) in which four amphiphilic peptide blocks with the propensity for helix formation were covalently attached to a heme moiety which served as the template molecule. A further application of the TASP concept which is currently being investigated is the construction of membrane channel models (56). In comparison with site directed mutagenesis for the mechanistic study of channel proteins (57, 58), engineering of channel proteins within

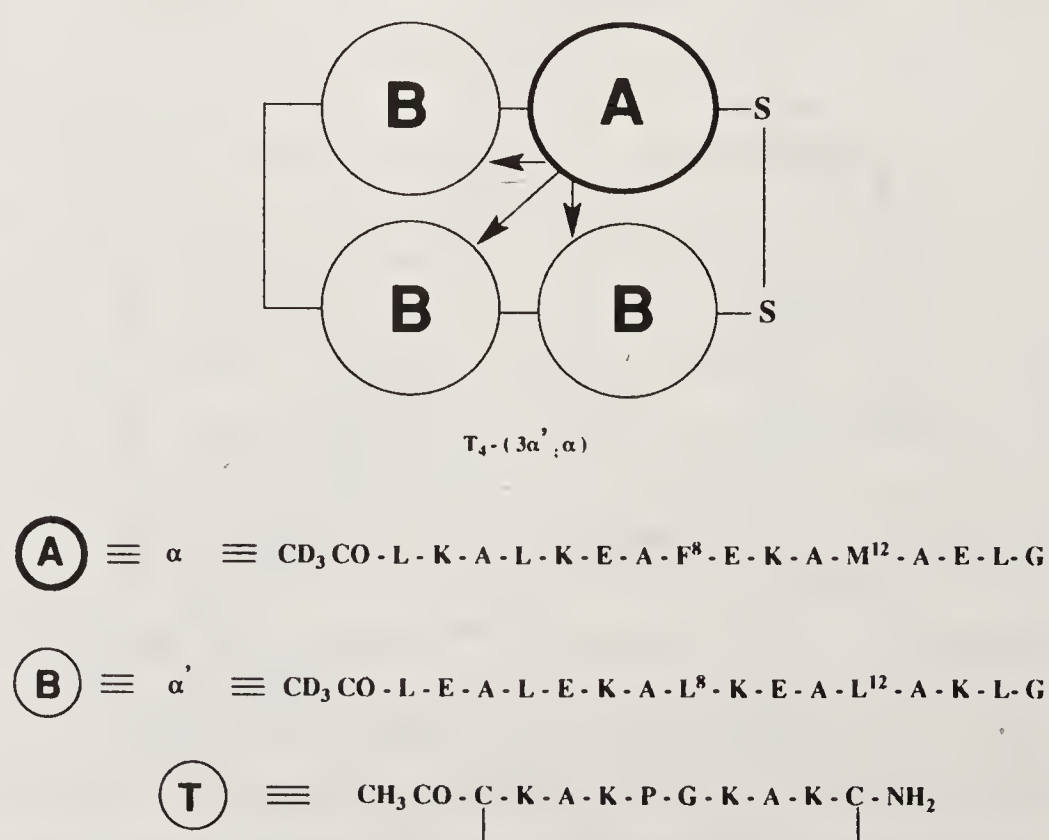


Figure 6. Diagram of TASP molecule $T_4-(3\alpha';\alpha)$, which contains two distinct peptide blocks attached to the cyclic carrier template molecule.

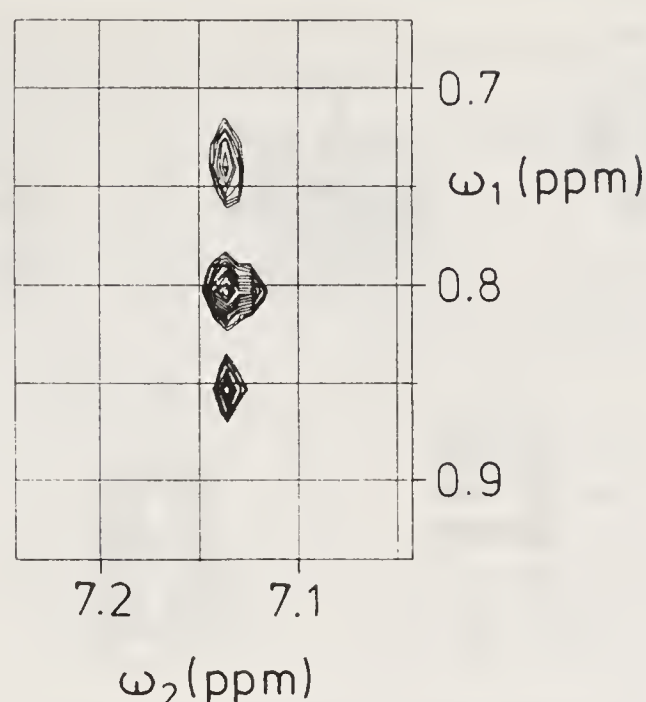


Figure 7. Aromatic (ω_2) versus methyl (ω_1) region of a NOESY spectrum of T_4 -($3\alpha'$; α), measured in H_2O , pH 2.7, at 47 °C with a mixing time of 150 ms. The only amino acid with an aromatic side chain is Phe 8 on the α -chain of T_4 -($3\alpha'$; α) and the chemical shifts around 7.15 ppm pertain to the phenyl ring protons of this residue. Contacts to three leucine methyl groups are evident (54).

the TASP framework offers distinct advantages (59) in that no ambiguity can arise in the distinction between altered function of the channel protein and problems which are associated with the cellular processing of the polypeptide or its assembly and folding in the membrane. Further, the so-designed channel protein can be studied *in vitro* in the absence of any other cellular process or component, and additionally, from a structural standpoint there is no ambiguity of the arrangement (parallel or antiparallel) of the channel-forming peptide blocks, since they are covalently attached to a carrier molecule.

Conclusions

In this chapter we have presented template assembled synthetic proteins, TASPs, as a conceptually new and widely applicable strategy for *de novo* protein design. Although a final proof of our proposed structures by X-ray analysis is still missing, the experimental data on the variety of TASP molecules synthesized to date have encouraged us to proceed with this approach. The challenge which lies immediately ahead of us is to improve upon our current synthetic strategies for TASP molecule construction and to achieve unequivocal proof of structure by means of nuclear magnetic resonance or X-ray analysis. Molecular recognition has been defined as a process that involves both binding and selection of substrate(s) by a receptor molecule, with the inclusion in the definition that the process as a whole should have a purpose, i.e. function (60). Artificial proteins (like their natural counterparts) possess immense inherent potential as receptor molecules for events of molecular recognition and have immense potential to fulfill the functionality as well. Certainly, as an ultimate goal in protein design, the incorporation of functional properties (catalytic and immunological activity, ion channel formation, or selective binding) into structurally defined novel proteins will be of utmost interest. It is our hope that the synthesis and

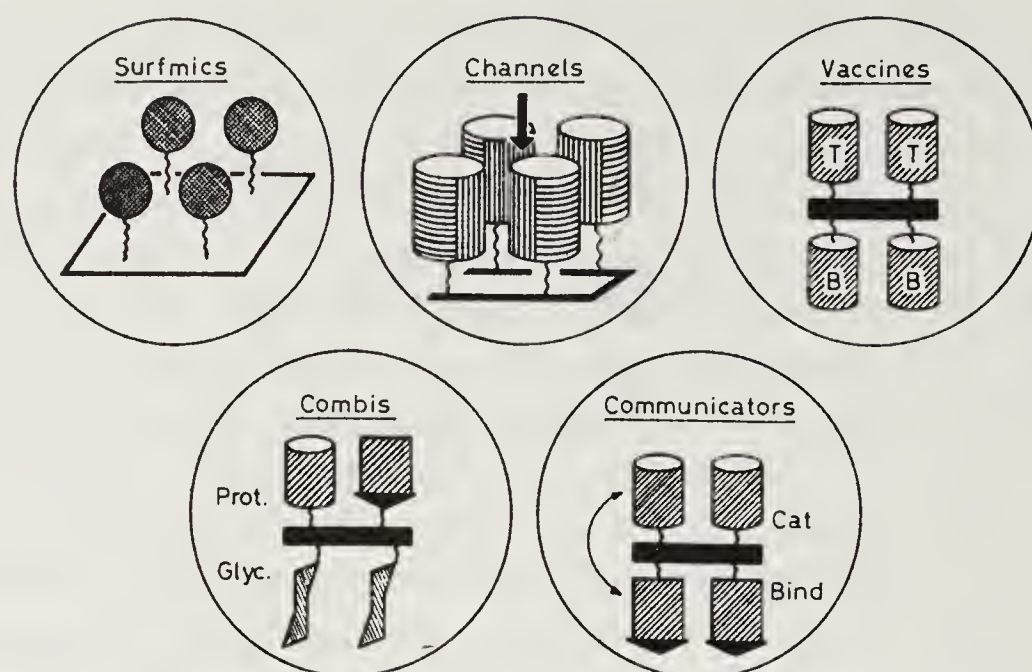


Figure 8. Schematic representation of possible “functional” TASP molecules. “Surfmics” would be surface mimetics in which the amino acid side chains of functionally relevant surface residues are assembled on a template matching the spatial requirements of the active sites of proteins. Ion specific “channels” could be constructed through the attachment of membrane-spanning helices to an appropriately designed template. Two domain TASPs could be “communicators”, containing both binding and catalytic sites, or synthetic “vaccines” containing both B and T cell epitopes. A “combis” might also be prepared in which peptidic and glycosidic elements are attached to a template molecule.

characterization of the possible TASP molecules shown schematically in figure 8 as well as other structures of biological importance will soon be achieved.

Acknowledgement

The authors gratefully acknowledge the financial support by the Swiss National Science Foundation, the Hoffmann-La Roche Research Foundation, and the Ciba Foundation. RIC gratefully acknowledges the postdoctoral training grant GM 13678-01 from the National Institutes of Health.

References

1. Royer, G. (1980) *Adv. Catal.* **29**, 197–227.
2. Lehn, J.-M. (1989) *Angew. Chem., Int. Ed. Engl.* **27**, 89–112.
3. Dugas, H. (1989) *Bioorganic Chemistry: A Chemical Approach to Enzyme Action*, 2 ed. (Springer-Verlag, New York).
4. Breslow R., ed. (1986) International Symposium on Bioorganic Chemistry, in *Ann. N. Y. Acad. Sci.* p. 471.
5. Davies, J. & Gassen, H.-G. (1983) *Angew. Chem., Int. Ed. Engl.* **22**, 13–31.
6. Botstein, D. & Shortle, D. (1985) *Science* **229**, 1193–1201.
7. Marshall, G., ed. (1988) *Peptides—Chemistry and Biology*, Proceedings of the 10th American Peptide Symposium, (Escom, Leiden).
8. Barany, G., Kneib-Cordonier, N. & Mullen, D. (1987) *Int. J. Pept. Protein Res.* **30**, 705–739.
9. Noren, C., Anthony-Cahill, S., Griffith, M. & Schultz, P. (1989) *Science* **244**, 182–188.
10. Tam, J. (1988) in *Peptides – Chemistry and Biology*, Proceedings of the 10th American Peptide Symposium, ed. Marshall, G. (Escom, Leiden) pp. 561–564.

11. Wüthrich, K. (1989) a. *Science* **243**, 45–53; b. Wüthrich, K. (1989) *Acc. Chem. Res.* **22**, 36–46.
12. Pace, N. (1990) *Trends Biochem. Sci.* **8**, 93–97.
13. Fersht, A. (1985) *Enzyme Structure and Mechanism*, 2nd ed. (WH Freeman, San Francisco).
14. Schimmel, P. (1990) *Acc. Chem. Res.* **22**, 233–234.
15. Knowles, J. (1987) *Science* **236**, 1252–1258.
16. DeGrado, W. (1989) *Adv. Protein Chem.* **39**, 59–124.
17. DeGrado, W., Wasserman Z. & Lear, J. (1989) *Science* **243**, 622–628.
18. Mutter, M. (1985) *Angew. Chem., Int. Ed. Engl.* **24**, 639–653.
19. Mutter, M. (1988) *Trends Biochem. Sci.* **13**, 261–265.
20. Mutter, M. & Altmann, K.-H. (1986) *The Construction of New Proteins*, *Die Angewandte Makromolekulare Chemie* 145/146, 211–229.
21. Mutter, M. & Vuilleumier, S. (1989) *Angew. Chem., Int. Ed. Engl.* **28**, 535–554.
22. Richardson, J. & Richardson, D. (1989) *Trends Biochem. Sci.* **14**, 304–309.
23. Gutte, B., Däumigen, M. & Wittschieber, E. (1979) *Nature (London)* **281**, 650–655.
24. Jaenicke, R., Gutte, B., Glatter, U., Strassburger, W. & Wollmer, A. (1980) *FEBS Lett.* **114**, 161–164.
25. Moser, R., Thomas, R. & Gutte, B. (1983) *FEBS Lett.* **157**, 247–251.
26. Lau, S., Taneja, A. & Hodges, R. (1984) *J. Biol. Chem.* **259**, 13253–12361.
27. Erickson, B., Daniels, S., Reddy, P. *et al.* (1986) *Current Communications in Molecular Biology, Computer Graphics, and Molecular Modeling*, Cold Spring Harbor Laboratory, New York, pp. 53–58.
28. Mutter, M., Altmann, K.-H. & Vorherr, T. (1986) *Z. Naturforsch.*, **41b**, 1315–1322.
29. Hodges, R., Semchuk, P., Taneja, A., Kay, C., Parker, J. & Mant, C. (1988) *Pept. Res.* **1**, 19–30.
30. Altmann, K.-H. & Mutter, M. (1990) *Int. J. Biochem.* **22**, 947.
31. Baldwin, R. (1989) *Trends Biochem. Sci.* **14**, 291–294.
32. Finkelstein, A. & Ptytsin, O. (1987) *Prog. Biophys. Mol. Biol.* **50**, 171–190.
33. Barany, G. & Merrifield, R. (1980) in *The Peptides – Analysis, Synthesis, Biology*, eds. Gross, E. & Meienhofer, J. (Academic Press, New York) Vol. 2, pp. 1–284.
34. Mutter, M., Altmann, E., Altmann, K. -H., Hersperger, R., Koziej, P., Nebel, K., Tuchscherer, G., Gremlich, H-U., & Müller, K. *et al.* (1988) *Helv. Chim. Acta* **71**, 835–847.
35. Mutter, M. & Tuchscherer, G. (1988) *Makromol. Chem., Rapid Commun.* **9**, 437–443.
36. Mutter, M., Hersperger, R., Gubernator, K. & Müller, K. (1989) *Proteins* **5**, 13–21.
37. Rivier, J., Miller, C., Andrews, J., Porter, J., Tuchscherer, G. & Mutter, M. (1990) in *Innovation & Perspectives in Solid Phase Peptide Synthesis*, ed. Epton, E. (SPPC, Birmingham, U.K.), pp. 39–50.
38. Rivier, J., Miller, C., Tuchscherer, G., Craig, A., Hernandez, J-F., Dykert, J., Raschdorf, F. & Mutter, M. (1990) *Proceedings of the 21st European Peptide Symposium*, in press.
39. Ernest, I., Vuilleumier, S., Fritz, H. & Mutter, M. (1990) *Tetrahedron Lett.* **31**, 4015–4018.
40. Ernest, I., Kalvoda, J., Rihs, G. & Mutter, M. (1990) *Tetrahedron Lett.* **31**, 4011–4014.
41. Wünsch, E. & Müller, E., eds. (1976) *Houben-Weyl, Methoden der Organischen Chemie*, (Thieme Verlag, Stuttgart), Vol. XV/2.
42. Kaiser, E., Mihara, H., Laforet, G. *et al.* (1989) *Science* **243**, 187–192.
43. Rink, H. (1987) *Tetrahedron Lett.* **28**, 3787–3790.
44. Kemp, D. (1981) *Biopolymers* **20**, 1793–1804.
45. Jakubke, H.-D. (1987) in *The Peptides: Analysis, Synthesis, Biology*, (Academic Press, New York), Vol. 9, pp. 103–165.
46. Wong, C-H, Chen, S-T., Hennen, W. *et al.* (1990) *J. Am. Chem. Soc.* **112**, 945–953.
47. Kemp, D. & Galakatos, N. (1986) *J. Org. Chem.* **51**, 1821–1829.
48. Kemp, D., Carey, R., Dewan, J., Galakatos, N., Kerkman, D. & Leung, S-L. (1989) *J. Org. Chem.* **54**, 1589–1603.
49. Kemp, D. & Carey, R., unpublished results.
50. Chou, K.-C., Maggoria, G., Némethy, G. & Scheraga, H. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4295–4299.
51. Shoemaker, K., Kim, P., York, E., Stewart, J. & Baldwin, R. (1987) *Nature (London)* **326**, 563–567.
52. Regan, L. & DeGrado, W. (1988) *Science* **241**, 976–978.
53. Mutter, M., Labhardt, A., Wyss, D., Rivier, J., Miller, C., Tuchscherer, G. & Carey, R., unpublished results.

54. Labhardt, A., Wyss, D. & Mutter, M., Unpublished results.
55. Sasaki, T. & Kaiser, E. (1989) *J. Am. Chem. Soc.* **111**, 380–381.
56. Montal, M., Montal, S. M. & Tomich, J. M. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6929–6933.
57. Imoto, K., Methfessel, C., Sakmann, B. *et al.* (1986) *Nature (London)* **374**, 670–674.
58. Imoto, K., Busch, C., Sakmann, B. *et al.* (1988) *Nature (London)* **385**, 645–648.
59. Montal, M. (1990) in *Ion Channels*, ed. Narahashi, T., (Plenum), Vol. 2, pp. 1–31.
60. Lehn, J-M. (1973) in *Struc. Bonding*, ed. Dunitz, J. *et al.* (Berlin) **16**, 1–71.

Peptide Models for Reverse Turns. The Role of Asparagine in the i Position of a β Turn

Josep Rizo, Madan M. Dhingra¹ and Lila M. Gierasch*

Department of Pharmacology, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235-9041, USA

Part I. Peptide Models for Reverse Turns

Introduction

Reverse turns are fundamental structural features of proteins and naturally occurring peptides where the peptide chain reverses its overall direction (1–3). The most common structures that bring about the chain reversal are β turns, which were first described by Geddes *et al.* (4) and defined by Venkatachalam (5), and involve four amino acid residues (numbered i to $i + 3$). β turns have been established as well defined structures that involve on average 25–30% of the residues in globular proteins (6, 7) and have been classified into different types according to the backbone dihedral angles of the central residues in the turn (1, 3, 5–8). Most frequently, the β turn structures are stabilized by a hydrogen bond between the NH of residue $i + 3$ and the CO of residue i . Although more than ten different types of β turns have been recognized, rationalization into seven different categories (I, I', II, II', VIa, VIb, VIII) appears to be the most sensitive classification (7, 8), and a new nomenclature, conveying intuitive conformational information, has recently been proposed by Wilmot and Thornton (8). Type I and Type II β turns are the most frequently observed turn conformations in proteins (7, 8) and will receive primary attention in this paper. Another possibility for the peptide chain reversal is the γ turn (9, 10), which involves three amino acid residues (numbered i to $i + 2$) and may include a hydrogen bond between the NH of residue $i + 2$ and the CO of residue i . In figure 1, Type I and Type II β turns, and a γ turn are represented with their standard torsion angles.

Turns are usually located at the protein surface (11), frequently connecting linear units of secondary structure such as α -helices and β -sheets, thus conferring on proteins their globularity (6). The residues involved in turns are predominantly polar (3, 6, 7, 11, 12). This and the fact that, in general, the NH and CO groups of the middle residues are not hydrogen bonded, while their side chains are pointing outwards,

¹ Present address: Tata Institute for Fundamental Research, Homi Bhabha Rd., Colaba, Bombay 400 005, India.

*For correspondence.

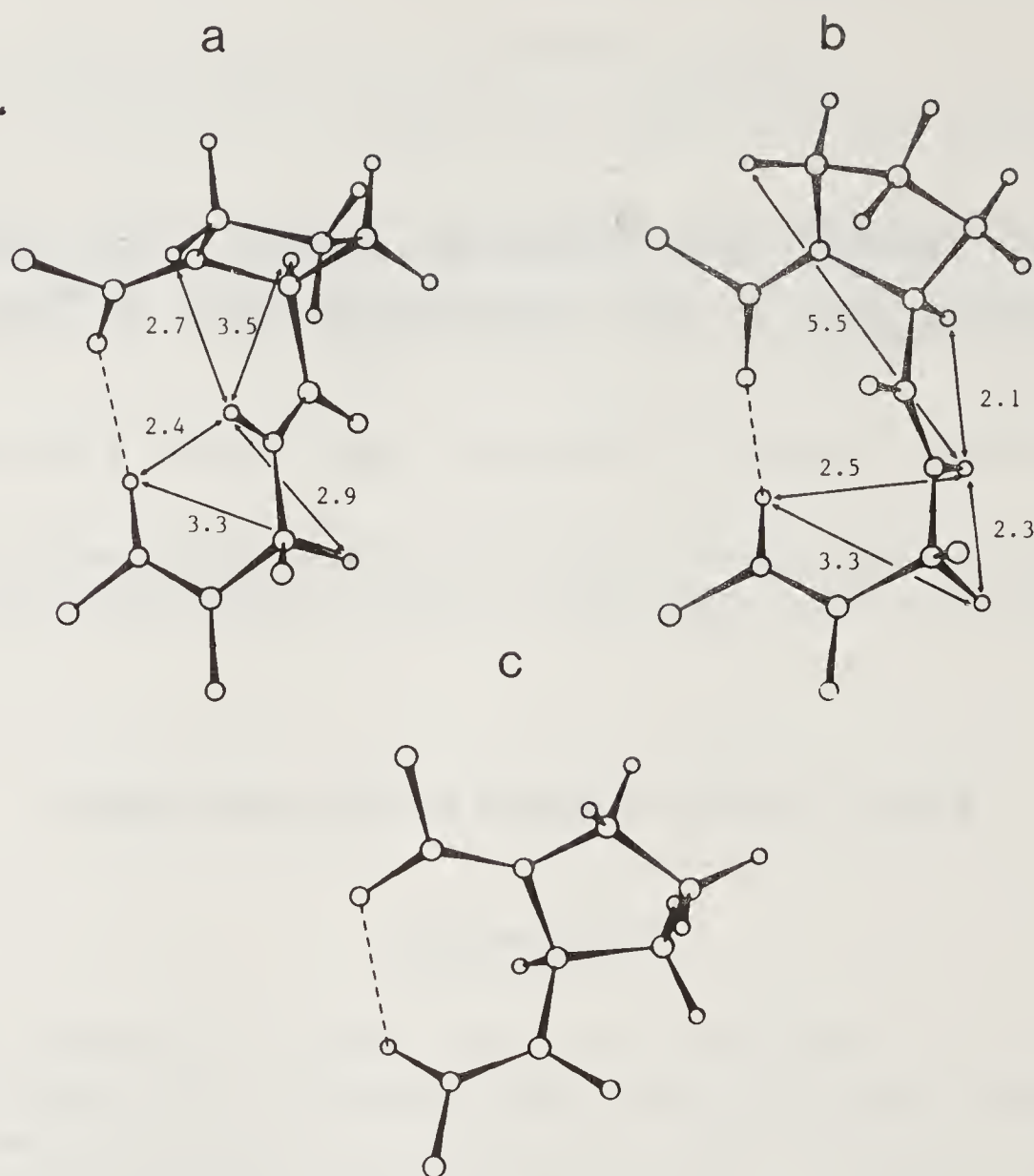


Figure 1. Perspective representation of three standard reverse-turn conformations with proline in the $i+1$ position: (a) Type I β turn ($\phi_{i+1} = -60^\circ$, $\psi_{i+1} = -30^\circ$, $\phi_{i+2} = -90^\circ$, $\psi_{i+2} = 0^\circ$); (b) Type II β turn ($\phi_{i+1} = -60^\circ$, $\psi_{i+1} = -120^\circ$, $\phi_{i+2} = 80^\circ$, $\psi_{i+2} = 0^\circ$); (c) γ turn ($\phi_{i+1} = -70^\circ$, $\psi_{i+1} = 70^\circ$). Hydrogen bonds are represented by dashed lines. Characteristic interproton distances in Type I and Type II β turns are indicated.

determine that turns are intrinsically polar structures, which is consistent with their surface location. The frequent occurrence of polar side chains within turns is also associated with their ability to interact with the unsatisfied amide groups stabilizing the turn conformation. These characteristics correlate with the observation that turns serve frequently as sites for receptor binding, antibody recognition and post-translational modification in proteins, although it is unclear whether this is due to a particular propensity to conformational recognition or to propitious surface localization (3). As described below, biological activity has often been related to turn formation in naturally occurring peptides, which supports the idea that the intrinsic compactness and the exposure of side chains in turns makes them ideal structures for receptor recognition.

The possible roles of turns in directing protein folding is a timely question that for the moment remains unanswered. On one hand, α -helices and β -sheets were recognized as the major stabilizing structures in proteins even before β -turns were established as well defined units of protein secondary structure (6), and turns have been implicated as passive structures that are formed due to long-range interactions within the polypeptide chain (13, 14). On the other hand, turns have been postulated to play an active role in protein folding (12, 15–18) and have been suggested to act

as a type of start/stop code at the end of helices and sheets (2). Statistical analysis of turns in proteins of known structure (6–8) has clearly shown not only that there are residues with high frequency of occurrence in turns, but also that there are strong positional preferences for some residues, and that the inclusion of turn type in the analysis clearly improves the quality of the prediction of turns in proteins based on these statistics (7, 8). The finding that, in general, almost one-third of the residues in a protein are involved in turns, the positional preferences for some residues, and the observation of significant populations of β -turn conformations in short linear peptides in aqueous solution (17, 18), argue for the possibility that turns may participate actively in protein folding. The occurrence of turns including mainly residues with low turn propensity points to a much more passive role for these turns. It seems reasonable to think, therefore, that turn formation is one of several factors that have to be taken into account when studying protein folding: some turns may direct the folding of a protein and/or stabilize its structure, while other turns may be more like accidental structures induced by other structural elements of the protein. Understanding of the rules for turn formation can help, therefore, to elucidate pathways of protein folding and to predict the preferred conformations of peptides and proteins.

Some insights into the factors that govern turn formation can be deduced from the statistical analysis of turns in proteins, and the inclusion of turns in protein secondary structure predictive methods has substantially improved their results (6–8). For instance, it is easy to rationalize the frequent occurrence of proline in the *i* position of β -turns by the constraint imposed by its cyclic structure on the ϕ dihedral angle. Statistical analysis of hydrogen bonding patterns in proteins has shown the ability of short hydrophilic side chains to establish specific hydrogen bonds with the peptidic backbone (19), and this ability can be related to their positional preferences in turns (6–8). However, several factors make it difficult to study in detail, within proteins, the nature of the interactions that may stabilize turns, the conformational characteristics that facilitate them, and their relative importance in energetic terms. The low resolution in the location of many side chains in the X-ray structures of proteins, the possible existence of interactions that may direct the folding of a protein or may influence its preferred conformation in solution, but may not be reflected in its crystal structure, and the size of a protein itself, are among the most important of these factors. Model peptides are more amenable to detailed conformational study, their sequences can be designed at will, and constraints can be imposed to induce particular structural features. A considerable amount of theoretical and experimental work has consequently been dedicated to the study of turns in peptide models (3). Rather than an exhaustive review of this work, in this paper we want to present an overview of what has been learned about turns from the study of model peptides and the implications of these studies (Part I). In Part II, we present some new results on the role of asparagine in the *i* position of a β turn obtained with a cyclic hexapeptide model.

General Considerations

Some characteristics of peptide models, which make them different from proteins, have to be taken into account in their conformational analysis (3). A first point,

intrinsic to the fact that small molecules are chosen as models, is that only short-range interactions determine their conformation. Also due to the small molecular size, most of the atoms in the peptide are exposed to the solvent and no distinction can be made between surface and interior residues. Consequently, their preferred conformations are strongly dependent on the environment. This is accentuated by the fact that they are often flexible molecules. In accordance with these characteristics, short linear peptides are generally unstructured in water and many studies are carried out in organic solvents with peptides constrained by cyclization. However, a few examples of somewhat structured linear peptides in aqueous solution have been described recently and are also discussed below. Although the flexibility of model peptides makes them difficult targets of conformational analysis, it also determines that they are good indicators of the relative stability of the conformational states that they can visit. Nevertheless, most successful studies with model peptides in solution have been carried out incorporating proline residues in the sequence to limit the conformational possibilities of the model.

Model peptides are usually obtained by chemical synthesis, and consequently it is easy to include non-natural amino acids and non-peptidic units in the model. Use of D-residues in peptide models has yielded experimental support for the predictions of the pioneering stereochemical calculations performed by Venkatachalam (5), later confirmed by more elaborate theoretical methods (20–22), on the influence of sequence chirality on the type of turn adopted. Homochiral (L–L) sequences in the central positions ($i + 1$ and $i + 2$) of β turns favor Type I β turns, while heterochiral (L–D) sequences favor Type II β turns; correspondingly, D–D and D–L sequences favor Type I' and II' β turns (the mirror images of Types I and II), respectively. These preferences are mainly explained by unfavorable steric interactions between the CO of residue $i + 1$ and the side chain of residue $i + 2$. Glycine can act as an L or a D residue, but most often does as a D residue, favoring Type II' or II β turns when it is located in the $i + 1$ or $i + 2$ position, respectively (3). This preference correlates with the high tendency of heterochiral sequences to form turns and with the frequent occurrence of glycine in the $i + 2$ position of Type II rather than Type I β turns in proteins (7), and points to a higher intrinsic stability of Type II and II' β turns with respect to Type I and I' turns. A stronger hydrogen bond closing the turn, due to a more favorable alignment of the participating NH and CO groups in Type II and II' β turns (as can be observed in figure 1), could contribute to this higher stability.

A variety of physical methods has been used to study turns in model peptides, including X-ray diffraction (23), nuclear magnetic resonance (NMR) (24), circular dichroism (CD) (25), infrared (IR) (26) and Raman spectroscopies (27), and molecular dynamics calculations (28, 29). X-ray diffraction is the technique of choice to study the conformation of peptides in crystals and can give very detailed structures, but care has to be taken to interpret the results as there is an interplay between intermolecular and intramolecular interactions (as shown in some examples below). NMR is the most useful method for studies in solution, especially when used in combination with molecular dynamics, and can give information about torsion angles, interproton distances, hydrogen bonding and dynamics. However, the time scale of NMR is slow compared to most of the conformational changes that can be expected in peptides and the observed parameters are, in general, averaged over the existing conformations. Solution techniques with fast time scales like CD, IR and Raman

spectroscopies offer, hence, complementary information, as the corresponding spectra will be the combination of the contributions from the different conformations. IR is especially useful to determine amide hydrogen bonding (30).

Model Dipeptides

The simplest models to study β turns are protected dipeptides of the form RCO-Xaa-Yaa-NHR'. Marraud and coworkers have worked extensively on these models and have shown their usefulness to reproduce β -turn conformations, both in crystals and in non-aqueous solutions (31–39). A good example of the results obtained with these models is offered by their work with sequences *t*-BuCO-L-Pro-Xaa-NHMe, with Xaa being an L or D aliphatic, aromatic or weakly polar residue (33). The bulky *t*-BuCO group is chosen to minimize the presence of *cis* conformer in its amide bond to proline. Following the rule mentioned above, these dipeptides assume, in solution, a Type I β -turn conformation for homochiral sequences (L-Xaa) and a Type II β turn when Xaa is a D-amino acid, as shown by IR and ^1H NMR. In contrast, all these dipeptides show a Type II β -turn structure in crystals, where a hydrogen bond between the Xaa NH and the Xaa CO of a neighboring molecule is observed in all cases. L–L sequences adopt Xaa ϕ torsion angles smaller than the standard values to accommodate the Type II β -turn conformation. This result shows that L–L sequences can adopt a Type II β -turn conformation if there are favorable interactions that afford some energy compensation (33), and underlines the influence of intermolecular interactions on the preferred conformations of model peptides. In a nice demonstration that short-range intramolecular interactions between polar side chains and the backbone amide groups also have their influence on conformational preferences, a Type I β -turn structure is found in the crystal of the corresponding dipeptide when L-Ser is the second residue in the sequence (34). This fact could be correlated with the frequent appearance of Ser in the $i + 2$ position of Type I β turns in proteins, in contrast with its very rare appearance in this position in Type II β turns (7). A weak hydrogen bond between Ser O γ and its own NH in the dipeptide crystals (34), which also seems to be retained in solution (35), could explain this tendency of serine residues in proteins and the different conformational behavior of serine with respect to apolar residues in the dipeptides. Even though this hydrogen bond has seldom been observed in protein crystals (19), favorable interactions are formed between the same polar groups (30). The same kind of interaction has also been found for *t*-BuCO-L-Pro-L-Asx-NHMe (Asx = Asp or Asn) in solution, where these peptides adopt mainly a Type I β -turn conformation (36). Asp residues show even stronger preference than Ser for Type I β turns, when they are located in the $i + 2$ position, in protein crystals (7). However, Asn residues are frequently found in the $i + 2$ position of both Type I and Type II β turns (7), which, by comparison with Asp, should be due to the NH₂ group of the amide side chain. This observation has been correlated with results obtained on model cyclic pentapeptides (40) as discussed below. Finally, it is noteworthy that NMR results obtained by Rao *et al.* for analogous dipeptide models in solution, show that the Pro-Gly and Pro-Aib (Aib, α -aminoisobutyric acid) sequences adopt Type II β -turn conformations, while a conformation with two consecutive γ turns is proposed for the Pro-Leu sequence (41).

Cyclized Turn Models

Two main classes of cyclized β -turn models have been studied, leading to similar conclusions to the ones obtained for the dipeptide models. One class consists of the general structure Cys–Xaa–Yaa–Cys, with both ends protected and a disulfide bridge closing the cycle (41–44). Most of these studies incorporate Pro as the first central residue (Xaa), while choices for the second residue include Aib, L-Ala, D-Ala, Gly, L-Leu, L-Val and D-Val. As shown by NMR, all these models contain β turns in solution, with the corresponding intramolecular hydrogen bond between Cys 4 NH and Cys 1 CO that seems to be favored by the disulfide bond (42). Homochiral sequences adopt Type I β turns, while heterochiral and Pro–Aib sequences prefer Type II β -turn conformations. Somewhat surprising is the observation of a Type I β turn for the Pro–Gly sequence, which could be favored by the constraints imposed in the model.

The other main class of cyclized β -turn models incorporates ϵ -aminocaproic acid (Aca) to link the carboxyl and amino ends of dipeptide sequences. These models have been studied in solution by NMR, CD and IR, and also by theoretical calculations (45–47). Studies carried out with L-Ala–Gly, L-Ala–D-Ala and L-Ala–L-Ala sequences showed the presence of Type II β turns for the two former ones (with some conformational averaging for L-Ala–Gly) and a Type I β -turn conformation for the last, and afforded good CD and IR patterns for the study of turn-containing peptides (45, 46). In another study, cyclo(D-Phe–L-Pro–Aca) was shown to adopt a Type II' β turn, following the rule of the heterochiral sequences, but with Pro in the $i + 2$ position in this case (47). The analogue L-Phe–L-Pro peptide exists as a single conformer with a *cis* amide bond between the two residues, while variable amounts of all-*trans* and one-*cis* conformers are observed for L-Pro–L-Phe and L-Pro–D-Phe sequences, the all-*trans* forms of these two peptides show Type I and Type II β -turn conformations, respectively. Of all these peptide models, only for the D-Phe–L-Pro sequence was convincing evidence found for a hydrogen bond closing the turn, but it is not clear if this is due to the constraints of the model. In this respect, it is noteworthy that related cyclic models of the general formula cyclo(NH-(CH₂)₄-CO-Gly–Pro–Xaa–Gly), with expected higher flexibility, present hydrogen-bonded β -turn conformations in DMSO-*d*₆ solution, as shown by NMR (48). The turn is Type II when the third residue is Gly and Type I when Xaa is Ser(O-*t*Bu). Interestingly, CD spectra of these peptides seem to indicate that these conformations are at least partially maintained in water solution.

Cyclic Pentapeptides

Cyclic pentapeptides are excellent models for reverse turns, as they can easily adopt a conformation consisting of fused β and γ turns, and they have been extensively studied both in crystals and in solution. A previous review covered the abundant literature through 1989 (40), and here we will only summarize the conformational possibilities of these peptide models, emphasizing some particular developments emerging from their study.

One of the first cyclic pentapeptides studied, cyclo(Gly–Pro–Gly–D-Ala–Pro), epitomizes the β, γ -turn conformation, with Pro 2–Gly 3 in the corner positions of

a Type II β turn and Pro 5 in the center of an inverse γ turn. Two hydrogen bonds, one between Gly 1 NH and D-Ala 4 CO and another between D-Ala 4 NH and Gly 1 CO, are found closing both turns. Both Gly residues adopt torsion angles generally preferred for D amino acids, so that the chirality of this conformation can be designated DLDDL. Indeed, incorporation of a D residue in positions 1 or 3 was entirely compatible with the β, γ -turn structure. The placement of an L residue in position 1 destabilizes this conformation, which is also present in cyclo(Ala-Pro-Gly-D-Phe-Pro) in solution, but with variable amounts of a one-*cis* (Ala-Pro) conformer depending on the solvent polarity. In cyclo(Gly-Pro-D-Phe-Ala-Pro), where an L residue has been placed in position 4, all four conformers with *cis* or *trans* Xaa-Pro bonds are observed. Cyclo(Gly-Pro-Ser-D-Ala-Pro) and cyclo(Gly-Pro-Ala-D-Phe-Pro), representatives of DLLDL sequences, were shown to adopt also a β, γ -turn conformation in solution, but the β turn lacked a strong hydrogen bond and the expected shift to a Type I turn could not be ascertained. Interestingly, the first peptide adopted a different conformation in crystals, where a Type II' β turn around Gly-Pro is observed. This conformation parallels the one described above for cyclo(D-Phe-L-Pro-Aca) and shows that Xaa-Pro sequences can be easily accommodated in Type II' β turns if the first residue can adopt torsion angles typical of D amino acids. In subsequent studies, cyclo(Gly-Pro-Ala-D-Phe-Pro) was compared to cyclo(D-Ala-Pro-Asn-Gly-Pro) using interproton distance determination by quantitative NOE measurement and molecular dynamics, and an equilibrium between Type I and Type II β, γ -turn conformations could be demonstrated for both peptides (40). The percentage of Type II conformer was 20% for the Pro-Ala sequence and 35% for the Pro-Asn sequence. This study showed that this kind of equilibrium could exist in many of the model peptides described above. Moreover, the increase in the Type II β turn proportion for the Pro-Asn sequence paralleled the aforementioned tendency of Asn to appear in the $i + 2$ position of Type II β turns in protein crystals (7). Note that quantitative interproton distance determination by NMR had been previously applied to cyclo(D-Phe-Pro-Gly-D-Ala-Pro) and the deduced conformation was perfectly comparable with the crystal structure of this peptide (49).

The work with cyclic pentapeptides has shown that γ turns can be persistent structural features in peptides and that a proline residue is particularly easy to accommodate in its central position. In fact, Ala, Val, Phe and D-Ala have also been shown to be compatible in the $i + 1$ position of γ turns. Furthermore, studies with only one proline-containing peptides have shown that a structure with two γ turns is another conformational possibility of cyclic pentapeptides.

Cyclic Hexapeptides

Cyclic hexapeptides have also been extensively used as models for turns, as they favor conformations with two fused β turns (3, 50), and work done with one of these models is presented in Part II. Many of these studies have been carried out with symmetric cyclo(Xaa-Pro-Yaa)₂ peptides in solution and have demonstrated their tendency to adopt this conformation. The general conclusions obtained with these peptides correlate with results described above for other models and are only briefly summarized. Proline residues are usually found in the $i + 1$ position of the turns, with Pro-Gly and Pro-D-Yaa sequences adopting Type II β turns. When Xaa

is glycine, Pro–Yaa sequences adopt, in general, Type I β turns, but a Type II' turn with Gly–Pro as central residues is observed when the bulky valyl side chain is in the Yaa position. This same conformation is observed for cyclo(D-Xaa–Pro–L-Yaa)₂ peptides, confirming the tendency of heterochiral sequences to adopt β turn conformations. Although all these conformations show C₂ symmetry in the NMR time scale, unfavorable transannular interactions between the carbonyls participating in the hydrogen bonds seem to exist, and deviations from this symmetry have been observed in crystals. This is not the case for cyclo(Ala–Pro–D-Phe)₂, for which a symmetric Type II β -turn structure is found in crystals, and has been the subject of a very significant theoretical study. Kitson and Hagler (51) have been able to reproduce the crystal structure of this peptide with molecular dynamics and energy minimization methods, and have found, in calculations carried out *in vacuo*, another conformation which is 8.5 kcal/mol lower in energy to the crystal conformation. These results underline the importance of intermolecular interactions in determining the preferred conformations of peptide models, and emphasize the ability of theoretical methods to reproduce experimental data. Finally, a recent study combining NMR and distance geometry calculations has shown that the cyclic bouvardin analogue, cyclo(Pro–MeTyr–Ala–MeTyr–MeTyr–D-Ala) adopts a conformation consisting of a Type II β turn around Pro 1–MeTyr 2 fused with a β turn around MeTyr 4–MeTyr 5 that include a *cis* amide bond between these two residues (52).

Linear Peptides

As mentioned in the introduction, short linear peptides are usually unstructured in aqueous solution. Many naturally occurring peptides have been proposed to adopt turn-containing structures that are implicated in their biological activity, but, in general this has been difficult to demonstrate by direct study of these peptides. However, design of analogues constrained by cyclization (forcing the turn conformation), and refinement of them following established structure–activity relationships, has led to the development of potent analogues that most likely mimic the structure of the native peptide bound to the corresponding receptor (3). A good example of this approach is gonadotropin-releasing hormone, where a Type II' β turn has been proposed to be an essential feature of the bioactive conformation (53) and has been found in our laboratory to be a common feature of very potent mono- and bicyclic-gonadotropin-releasing hormone antagonists (Bienstock, R. J., Rizo, J. & Gierasch, L. M., unpublished results). Other approaches used to constrain analogues of naturally occurring peptides are the introduction of bulky groups in the side chains and the incorporation of lactam bridges between residues; studies on enkephalin are a good example of these approaches (54).

Despite the lack of structure that many linear peptides show in aqueous solution, some examples have been reported recently where a sizable population of turn-structures has been demonstrated. Perhaps the most significant of these studies is the one reported by Wright and coworkers where different amounts of β -turn conformation are observed in linear penta- and hexapeptides derived from the sequence Tyr–Pro–Tyr–Asp in aqueous solution (17, 18). In particular, these authors have found, for pentapeptides of general formula Tyr–Pro–Xaa–Asp–Val, a correlation between the β -turn populations observed and β -turn probabilities

obtained from protein crystals statistics. These results show that, in polypeptide chains, some sequences can frequently visit turn conformations, which could have a major effect in protein folding pathways (17, 18). In another study directed to assign secondary structure elements in proteins with anti-peptide antibodies, a fifteen residue peptide with a central sequence Asp-Pro-Gly-Gln has been shown to adopt at least 50% of the time a β -turn conformation connecting two strands of antiparallel β -sheet structure (55). It has also been shown by Raman spectroscopy that poly(Val-Pro-Gly-Val-Gly) in aqueous solution adopts a similar conformation to that of cyclo(Val-Pro-Gly-Val-Gly)₃, which is a model for protein Type II β turns based on a sequence from elastin (56). A Type II β turn encompassing residues Lys 7-Pro 8-Gly 9-Gln 10 is observed in the α -factor from *Saccharomyces cerevisiae* both in DMSO and water solution, and the ability to assume this conformational feature has been correlated with biological activity in analogues of this tridecapeptide (57). Finally, recent results obtained in our laboratory on isolated peptides from the P22 bacteriophage tailspike protein (58) have been correlated with the model that has been proposed for its folding. Several mutants of this protein have been found to be temperature-sensitive for folding by King and coworkers (59). Wildtype decapeptide and dodecapeptide sequences corresponding to a region of the protein that includes one of these mutations have a three- to five-fold higher tendency to adopt turn conformations than analogous peptides corresponding to the mutant sequences (58). All of these results argue for an active role of turn formation in the mechanisms that lead to the folding of a protein and underline the usefulness of peptide models to study these mechanisms.

Part II. A Cyclic Hexapeptide Model for Turns with Asparagine in the i Position

Introduction

In part I we have described what kind of information on turn preferences in peptides and proteins has been obtained through the study of model peptides and what questions can be addressed. In this part we show how this information can be obtained on a cyclic hexapeptide model of a protein β turn, through the combined use of high field nuclear magnetic resonance and molecular dynamics calculations. We had particular interest in determining the roles of asparagine in turn formation due to its very high tendency to appear in protein β turns (6–8). As mentioned above, we reported recently a study on the influence of asparagine on the β -turn type when this residue appears in the $i + 2$ position, using a cyclic pentapeptide model (40). In the present study we explore the function of asparagine when it is located in the i position of a β turn.

Asparagine is an amino acid residue with very particular properties. The amide group in its side chain mimics the backbone peptide groups and can interact with them, both as a hydrogen bond donor and as an acceptor. The short length of this side chain limits its conformational possibilities and favors the formation of short range interactions, as has been shown by statistical analysis of intramolecular

hydrogen bonds in protein crystals (19). The most frequent of these interactions is a hydrogen bond between the Asn O^δ in residue n and the backbone NH of residue $n + 2$. This hydrogen bond implies the formation of a ten-membered ring in a β -turn like structure, referred to as an Asx turn (1, 60), where the Asn O^δ plays the role of the backbone O in position i . This ability of the Asn O^δ to mimic the backbone O of the residue preceding asparagine also leads to the frequent occurrence of this residue as an N-terminal cap of α -helices (61).

Asparagine is the amino acid residue which appears most frequently in the i position of Type I β turns in proteins of known structure, followed very closely by Asp and Ser (7). The same interaction described above, a hydrogen bond between the side chain of these residues and the NH of residue $i + 2$ (figure 2), seems to make these turns very stable (1), and is found in 57% of the Type I β turns containing Asn, Asp and Ser in this position (7). In contrast, no preference for these residues in the i position is observed in Type II β turns, where the NH of residue $i + 2$ is no longer pointing towards the i side chain (7). Although some conformational studies of the Asx turn with model tripeptides have been reported (37–39), no detailed conformational analysis has been carried out on the interactions that can occur between the asparagine side chain and the peptide backbone when this residue appears in the i position of a β turn. To address these questions, we chose the cyclic hexapeptide cyclo(Asn–Pro–Ala–Leu–Pro–Gly) as a model. The sequence Asn–Pro–Ala–Leu corresponds to the segment 148–151 of dihydrofolate reductase (*L. casei*), which assumes in the protein crystal the structural pattern described above: a hydrogen bonded Type I β turn, stabilized by an Asx turn (figure 2). According to the experience accumulated through the study of cyclic hexapeptides, we expected the Pro–Gly sequence to assume a Type II β -turn conformation that could induce the formation of a β turn on the other side of the ring. We have studied the model peptide cyclo(Asn–Pro–Ala–Leu–Pro–Gly) by high field NMR, and found that this turn is adopted by the peptide in its favored conformation in chloroform solution (including

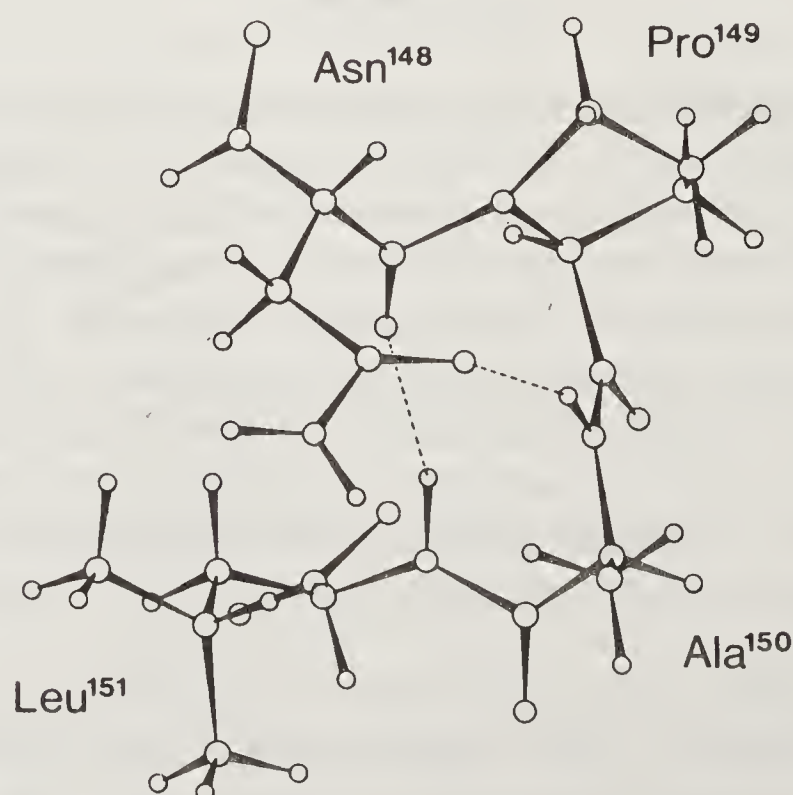


Figure 2. Conformation of the Asn 148–Pro–Ala–Leu 151 segment of dihydrofolate reductase (*L. casei*), containing a Type I β turn and an Asx turn stabilized by two hydrogen bonds (dashed lines).

14% DMSO- d_6); this conformer also contains a hydrogen bond between the Asn 1 O $^\delta$ and the Ala 3 NH. Measurement of interproton distances through nuclear Overhauser effect (NOE) quantitation, combined with molecular dynamics simulations and energy minimization, shows that several rotamers of the Asn side chain, which could exist in rapid equilibrium in the NMR time scale, allow this interaction to be established. Furthermore, our results show that an interaction between the Asn 1 O $^\delta$ and the Leu 4 NH can also exist, which could induce and stabilize the formation of the β turn. Finally, a conformational change is observed in DMSO- d_6 solution, where the Asn side chain appears to be hydrogen bonding to one of the backbone carbonyls of the molecule.

Experimental Methods

Synthesis

The synthesis of the cyclic hexapeptide was performed by standard methods described elsewhere (62). Amino acids were purchased from Bachem or Peptides International. Couplings were carried out by the mixed anhydride method, using isobutylchloroformate as activating agent, and N-methylmorpholine as a base. The *t*-butyloxycarbonyl (Boc) group was used to protect the amino-termini, and was removed with a solution of HCl in diethyl ether. Benzyl (Bz) esters were used for C-terminal protection, and were cleaved by hydrogenolysis in *t*-butanol with 10% Pd/C as a catalyst. Cyclization was performed at high dilution in pyridine, after hydrogenolysis of the linear form Boc-Asn-Pro-Ala-Leu-Pro-Gly-OBz, C-terminal activation with *p*-nitrophenol and removal of the Boc group.

NMR Spectroscopy

Most NMR experiments were carried out with a 15 mM solution of cyclo-(Asn-Pro-Ala-Leu-Pro-Gly) in CDCl $_3$, containing 14% DMSO- d_6 to keep the peptide in solution (total volume 0.75 ml). Some NMR data were acquired on a 5 mM sample of the cyclic hexapeptide in DMSO- d_6 . The solvent titration was performed mixing different volumes of 5 mM solutions of the cyclic hexapeptide in DMSO- d_6 and in CDCl $_3$ (14% DMSO- d_6). NMR tubes were purchased from Wilmad Glass Co. (5 mm, 526 pp), and deuterated solvents were "100 atom% D", from Sigma Chemical Co.

^1H and ^{13}C -NMR spectra were recorded on Varian VXR-500 and Bruker AC-250 spectrometers, respectively. Two-dimensional total correlation (TOCSY) (63), nuclear Overhauser (NOESY) (64–66), and rotating-frame Overhauser (ROESY) (67, 68) spectroscopies were used to assign the resonances of the proton spectrum of the cyclic hexapeptide, and to obtain quantitative NOE data. TOCSY and NOESY spectra in CDCl $_3$ (14% DMSO- d_6) were recorded at 278 K. ROESY spectra in CDCl $_3$ (DMSO- d_6), and TOCSY and NOESY spectra in DMSO- d_6 , were obtained at 298 K. All of them were recorded in the phase sensitive mode (69), with 256 t_1 values and 1024 complex points per FID. The number of scans per t_1 value varied between 16 and 64, and the spectral width was 6000 Hz. A relaxation delay of 1 s was used in general, except for the quantitative NOESY and ROESY spectra, where it was set to 2 s. TOCSY spectra were obtained with a 5 kHz spin-lock field, 75 ms mixing time and a 1 ms trim pulse. The mixing time for NOESY spectra was 0.5 s, and this was

lowered to 100 ms to obtain a quantitative data set. For ROESY spectra, 100, 150, and 250 ms mixing times were used. A low spin-locking field (2 kHz) (70) and short (30°) mixing pulses (68) were used to minimize the contributions of J cross-peaks, and two hard 90° pulses were employed at both sides of the spin lock to achieve resonance-offset compensation.

Two-dimensional NMR data was processed on a SUN 4/260 computer using software developed by Dennis Hare (71). Gaussian apodization was used, in general, for both time domains, except for quantitative NOESY and ROESY spectra, where a 90° -shifted sine bell was employed. The spectrum corresponding to the first t_1 value was multiplied by 0.5 before the Fourier transform along the t_1 axis to eliminate noise in this dimension (72). Zero-filling was applied in the t_1 dimension to obtain a matrix of 1024×1024 real points. Interproton distances were calculated from one NOESY (100 ms mixing time) and two ROESY (100 and 150 ms mixing times) spectra. Fourth order polynomial base-line correction was applied in both dimensions in these spectra, and volume integrals were obtained on both sides of the diagonal in non-symmetrized data. The NOE between the two Pro 2 H^δ protons was used as a reference, assuming an interproton distance of 1.79 Å.

Molecular Dynamics

Molecular dynamics simulations and energy minimizations were performed on SUN 4/260 or Silicon Graphics Personal Iris computers, using the program DISCOVER (73). The molecular modeling programs MOLEDT and INSIGHT II (73) were used to build initial structures and to analyze the results of the calculations. A total of 6 molecular dynamics trajectories at different temperatures, and with different starting structures (as explained in the text), were calculated. All the simulations were performed *in vacuo*, using a "leapfrog" algorithm with a time step of 1 fs for 0.5 fs (for high temperature calculations) to integrate the equations of motion: Instantaneous structures were taken at 0.1 ps intervals along all the trajectories, and 10% of these structures (at 1 ps intervals) were energy minimized. Energies, torsion angles, and hydrogen bonding patterns were extracted from all these structures and analysed. Some interproton distances measured by NMR (see table 3 below) were used as constraints in one of the simulations. To impose the constraints, a skewed biharmonic potential with a $30 \text{ kcal}/\text{\AA}^2 \text{ mol}$ force constant was added to the force field. This potential is null between a specified distance range, which was chosen according to the estimated errors in the NMR distances (table 3 below). The distances between Asn 1 H^α and Pro 2 C^δ , and between Leu 4 H^α and Pro 5 C^δ , rather than the interproton H^α/H^δ distances, were constrained to keep both H^δ protons of each proline close to the H^α proton of the preceding residue. The shortest Xaa(i) $H^\alpha/\text{Pro}(i+1) C^\delta$ distance was found to be 2.6 Å by computer graphics, and, accordingly, a 2.5–2.7 Å range was used in the constraints.

Results

Conformational Analysis by NMR in CDCl_3 (14% $\text{DMSO}-d_6$)

The 500 MHz ^1H -NMR spectrum of cyclo(Asn–Pro–Ala–Leu–Pro–Gly) in CDCl_3 (14% $\text{DMSO}-d_6$) solution is shown in figure 3. The assignment of all the resonances

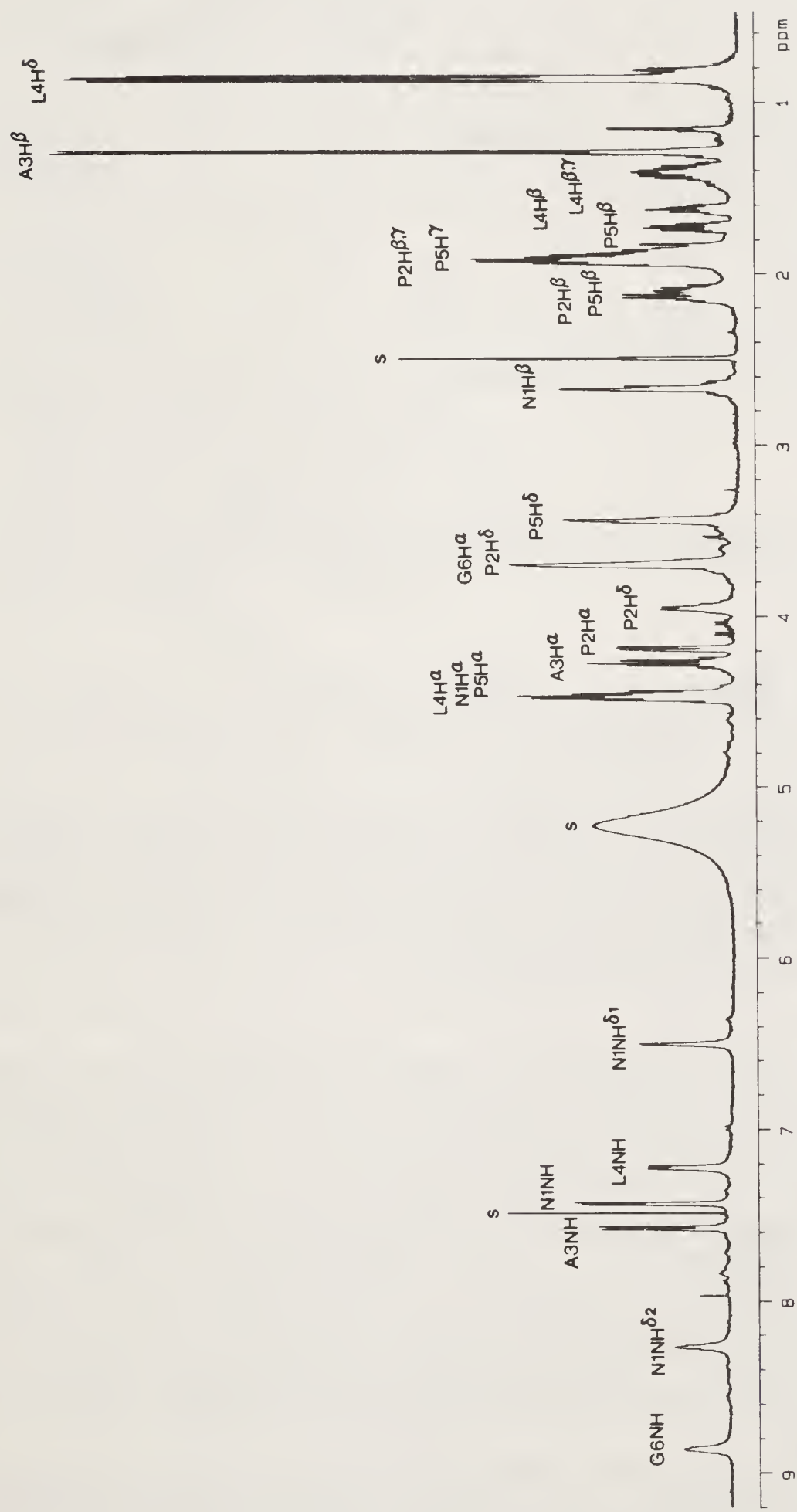


Figure 3. One-dimensional 500 MHz ¹H-NMR spectrum of cyclo(Asn-Pro-Ala-Leu-Pro-Gly) in CDCl₃ (14% DMSO-d₆). The assignment of all the proton resonances of the molecule is indicated. Solvent peaks (CHCl₃, DMSO and water) are labelled with an s.

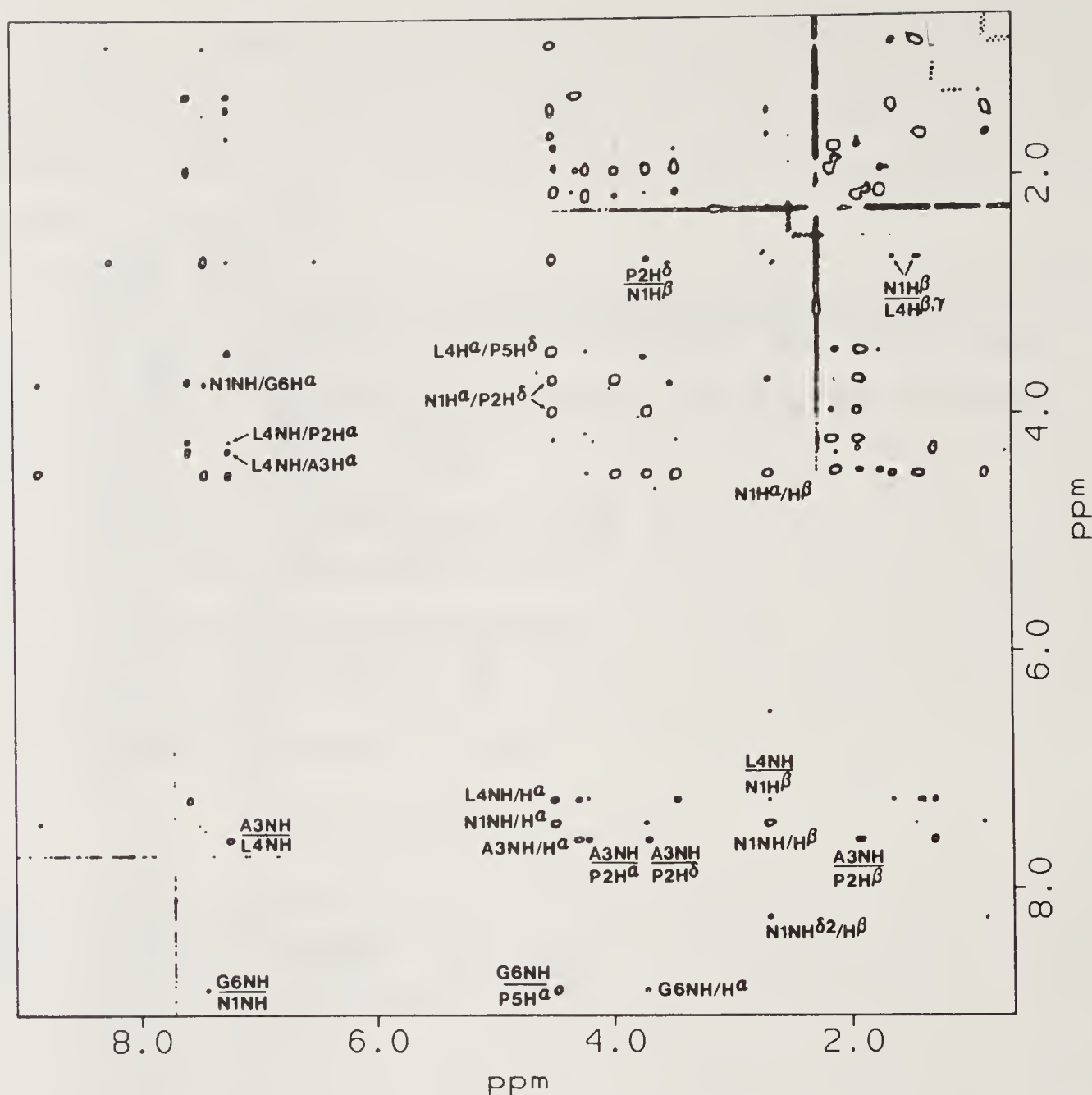


Figure 4. Contour map of a 500 MHz ^1H phase-sensitive ROESY spectrum of cyclo(Asn-Pro-Ala-Leu-Pro-Gly) in CDCl_3 (14% DMSO-d_6) at 298 K (0.25 s mixing time). The data have been symmetrized about the diagonal. Cross-peak assignments are indicated with the involved protons separated by a slash or a dash, and the downfield resonance listed first.

in the spectrum is easily accomplished through the combined use of the TOCSY and NOESY spectra at 278 K. Due to the small size of the molecule, and the high field used in this study, the temperature had to be lowered to observe sizeable NOEs. The ROESY spectrum at room temperature (figure 4) gives essentially the same NOE pattern, but with stronger intensities, thus facilitating the conformational analysis. In table 1 are summarized the most important NMR parameters that are used, together with the NOE information, to determine the overall conformation of the molecule. Refinement of this structure was accomplished by interproton distance measurements, using quantitative NOEs and molecular dynamics simulations, which are described in the next sections.

In the ^1H -NMR spectrum of cyclo(Asn-Pro-Ala-Leu-Pro-Gly) (figure 3) it can be observed that the molecule exists as a single conformer in CDCl_3 (14% DMSO-d_6) solution. The small differences in the chemical shifts of the β and γ carbons of both Pro 2 and Pro 5 (74), as well as the strong NOEs between Asn 1 H^α and Pro 2 H^δ protons, and between the Leu 4 H^α and Pro 5 H^δ protons, indicate that the major form of the peptide in these conditions is the all-*trans* conformer. The low temperature dependence of the chemical shift of the Asn 1 and Leu 4 amide protons (table 1)

Table 1. NMR parameters for cyclo(Asn–Pro–Ala–Leu–Pro–Gly)^a

Parameters	Asn	Pro	Ala	Leu	Pro	Gly
$\delta_{\text{NH}}(\text{ppm})$ NH	7.43		7.58	7.22		8.87
NH ₂	8.27 ^b					
	6.50 ^c					
$\Delta\delta/\Delta T(\text{ppb/K})$ NH	–1.2		–1.6	0.4		–4.3
NH ₂	–7.0 ^b					
	–2.3 ^c					
$\delta\text{H}^\alpha(\text{ppm})$	4.47	4.19	4.27	4.49	4.44	3.70 ^d
$\delta\text{H}^\beta(\text{ppm})$	2.67 ^d	2.13	1.28	1.62	2.09	
		1.93		1.40	1.73	
$^3J_{\text{HN}\alpha}(\text{Hz})$	4.6		8.3	7.1		^e
$\delta^{13}\text{C}\beta(\text{ppm})$	37.1	29.0	18.2	41.0	29.0	
$\delta^{13}\text{C}\gamma(\text{ppm})$		24.1 ^f		24.0 ^f	24.1 ^f	

^a Estimated absolute errors: ¹H chemical shifts 0.02 ppm; ¹³C chemical shifts 0.1 ppm, coupling constants 0.3 Hz

^b NH ^{δ^2} (*trans* to 0 ^{δ} in the amide bond)

^c NH ^{δ^1} (*cis* to 0 ^{δ} in the amide bond)

^d Equivalent methylene protons

^e Both < 5.5 Hz, as estimated from non-resolved Gly 6 NH resonance

^f Interchangeable

indicates that they are sequestered from the solvent and may be participating in intramolecular hydrogen bonds. This information, together with the observation of NOEs between Ala 3 NH and Leu 4 NH, and between Gly 6 NH and Asn 1 NH, argues that the expected conformation, consisting of two fused β turns with Pro 2–Ala 3 and Pro 5–Gly 6 in the corner positions, is adopted by the peptide in CDCl₃ (14% DMSO-*d*₆). Other NOE connectivities also agree qualitatively with the NOE patterns expected for β turns (40, 41, 75, and figure 1). The small intensities of the NOEs between Ala 3 H ^{α} and Leu 4 NH, and between the Gly 6 H ^{α} protons and Asn 1 NH, correlate with the presence of the two β turns. Furthermore, an NOE between Pro 2 H ^{α} and Leu 4 NH, also characteristic of β turns, is observed, while the presence of an NOE between Pro 5 H ^{α} and Asn 1 NH cannot be assessed because of overlapping with the intraresidue Asn 1 H ^{α} /NH connectivity. Given that the distance between the H ^{α} proton of residue *i* + 1 and the NH of residue *i* + 3 in standard Type I and Type II β turns is 3.6 and 3.3 Å, respectively (75), we can assume that NOEs between protons up to about 3.5 Å apart can be observed in our ROESY data.

The distance between the H ^{α} proton of residue *i* + 1 and the amide proton of residue *i* + 2 is the most sensitive parameter to distinguish between Type I and Type II β turns, but care must be taken when interpreting the NOE data due to the possibility of conformational averaging between both turn types (40). The strong NOE between Pro 5 H ^{α} and Gly 6 NH suggests that the Pro 5–Gly 6 turn is Type II, as could be expected, but some Type I β -turn population cannot be discounted. In the Pro 2–Ala 3 region, a weak NOE is observed between Pro 2 H ^{α} and Ala 3 NH. Taking into account that a small percentage of Type II β -turn conformation can result in a sizable NOE (40), we can propose that the β turn around Pro 2–Ala 3 is predominantly Type I. The observation of an NOE between Ala 3 NH and one of the Pro 2 H ^{δ} protons agrees with this assumption, while the absence of any

connectivity between Gly 6 NH and the Pro 5 H $^{\delta}$ protons correlates with the presence of a Type II β turn around Pro 5–Gly 6.

The temperature coefficient of Ala 3 NH (table 1) is also low, suggesting that this amide proton is participating in a hydrogen bond. The expected interaction between Ala 3 NH and the Asn 1 side chain appears as the most likely explanation for this observation, but the conformational analysis of the rotameric state of the Asn 1 side chain is hindered by the overlapping between the two H $^{\beta}$ protons. This overlapping precludes the differentiation between NOE connectivities to the two methylene protons and the measurement of H $^{\alpha}$ /H $^{\beta}$ coupling constants. Nevertheless, the presence of interactions from the Asn 1 methylene group to the Leu 4 amide and side chain protons indicates that the Asn 1 side chain is located above the hexapeptide ring. Model building shows very clearly that this has to be the case to form the proposed hydrogen bond closing an Asx turn. The model also suggests that the upfield Asn 1 NH $^{\delta}$ proton (NH $^{\delta 1}$, *cis* to O $^{\delta}$ in the amide bond) could be somewhat buried in the middle of the molecule, which agrees with the relatively low temperature coefficient of this proton (table 1).

Solvent Titration with DMSO- d_6

The difference between the chemical shifts of the amide protons in DMSO- d_6 (a good hydrogen bond acceptor) and CDCl $_3$ solution is also indicative of the degree of exposure of these protons to the environment, and hence their possible implication in intramolecular hydrogen bonds, assuming that there is no conformational change. A solvent titration, in which the percentage of DMSO- d_6 in the solvent system was varied from 14% to 100%, was carried out. The results of this experiment (figure 5) are in clear contrast with results of similar titrations reported for other model peptides (e.g., ref. 76), where most of the change in chemical shifts occurs for low amounts of DMSO- d_6 (up to ca. 20–40%), and are strongly indicative of a conformational change. The absolute change in chemical shifts in the solvent titration and the temperature

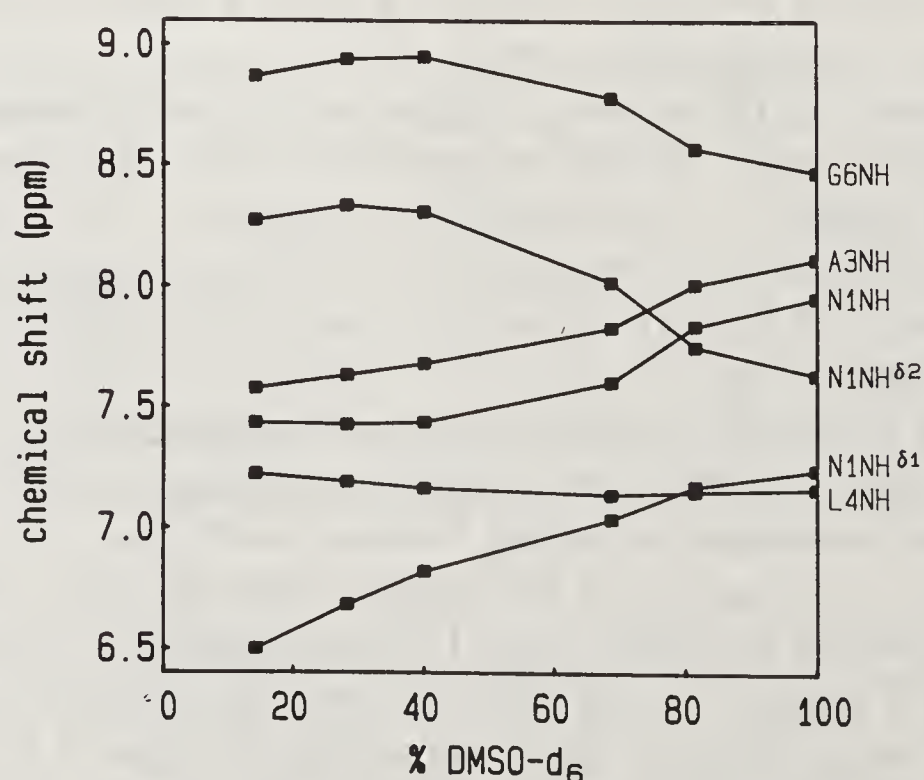


Figure 5. Dependence of the chemical shifts of the amide protons of cyclo(Asn–Pro–Ala–Leu–Pro–Gly) on the DMSO- d_6 content (v/v) in different CDCl $_3$ /DMSO- d_6 solvent mixtures.

Table 2. Difference between the chemical shifts of the amide protons of cyclo(Asn-Pro-Ala-Leu-Pro-Gly) in DMSO-d₆ and CDCl₃ (14% DMSO-d₆), and temperature coefficients in DMSO-d₆.

	Asn 1 NH	Asn 1 NH ^{δ1}	Asn 1 NH ^{δ2}	Ala 3 NH	Leu 4 NH	Gly 6 NH
$\Delta\delta(\text{ppm})^a$	0.52	0.74	-0.66	0.53	-0.06	-0.41
$\Delta\delta/\Delta T$ (ppb/K)	-8.9	-4.9	-1.3	-5.6	-0.4	0.8

^a(DMSO-d₆)- δ [CDCl₃(14% DMSO-d₆)]

coefficients obtained in DMSO-d₆ (table 2), compared with the CDCl₃ (14% DMSO-d₆) coefficients (table 1), give clear support to this conclusion. The Leu 4 NH is the only amide proton that remains sequestered in both solvent systems. In the first part of the titration curve, the Asn 1 and Ala 3 amide protons are quite insensitive to the variation of DMSO-d₆ contents in the solution, as corresponds to non-exposed protons, but a clear increase in their chemical shifts is observed between 70% and 80% DMSO-d₆. These protons seem clearly exposed in 100% DMSO-d₆ (table 2). The chemical shifts of Gly 6 NH and the downfield Asn 1 NH^δ (NH^{δ2}) protons reach a maximum for ca. 30–40% DMSO-d₆, but a decrease in these chemical shifts is observed for higher DMSO-d₆ contents, which is clearly accentuated between 70% and 80% DMSO-d₆. These protons seem to be hindered from the solvent in 100% DMSO-d₆ (table 2), which can explain the unusual decrease of the chemical shift in this solvent with respect to CDCl₃ (14% DMSO-d₆), where both protons are exposed and can interact with the solvent hydrogen bond acceptor. It has to be noted that some of the temperature dependence curves in DMSO-d₆ are not completely linear, which indicates that changes in the populations of conformations are occurring when the temperature is varied.

A preliminary assessment of the conformational change occurring at high percentages of DMSO-d₆ was attempted, but no detailed analysis was pursued in this study. The TOCSY and NOESY spectra of cyclo(Asn-Pro-Ala-Leu-Pro-Gly) in DMSO-d₆ allow the assignment of all the proton resonances of the molecule. The high viscosity of DMSO-d₆ lowers the correlation time enough to obtain good NOESY data at room temperature. The NOE pattern observed is quite similar to the one observed in CDCl₃ (14% DMSO-d₆), but many fewer contacts between the Asn 1 side chain and the Leu 4 protons are observed. An increase in the relative intensity of the Pro 2 H^α/Ala 3 NH NOE suggests that Type II β -turn conformations around Pro 2–Ala 3 are more frequently visited than in CDCl₃ (14% DMSO-d₆). This could be due to the lack of an Asn 1 O^δ/Ala 3 NH hydrogen bond which stabilizes the Type I β turn. Although no clearcut structure was drawn from these data, model building suggests that Asn 1 NH^{δ2} (*trans* to the O^δ in the amide bond) could be hydrogen bonding to Gly 6 O, disrupting the Pro 5–Gly 6 β turn, and a γ turn around Pro 5 could be favored. A clear conclusion is that the presence of a good hydrogen bond acceptor such as DMSO-d₆ can disrupt the structure with two β turns and an Asx turn proposed above – leading to conformations with a higher number of exposed amide protons – but only at high concentrations of DMSO-d₆. More important, the Asn 1 side chain appears to be interacting with one of the

backbone carbonyls of the peptide and participating actively in the conformational change.

Measurement of Interproton Distances

Interproton distance measurement by quantitative NOE analysis, combined with the use of molecular dynamics calculations, is a powerful method to refine structures of model peptides that have been roughly delineated by NMR. The initial buildup rate of an NOE cross peak between two proton resonances can be related to the inverse of the sixth power of the distance between these two protons and used to measure it, provided that an interaction between two protons of the molecule whose interproton distance is known can be used as a reference (64, 66). NOEs measured at short mixing times, divided by the corresponding mixing time, can be used to obtain a good approximation of the initial buildup rate. To obtain interproton distances for cyclo(Asn-Pro-Ala-Leu-Pro-Gly) in CDCl₃ (14% DMSO-d₆) solution, we recorded a NOESY spectrum with a mixing time of 100 ms, and two ROESY spectra with 100 ms and 150 ms mixing times. Interproton distances were calculated, independently, from the cross peak intensities on each side of the diagonal for each spectra. The NOE between the two H^δ protons of Pro 2, whose interproton distance was assumed to be 1.79 Å, was used as a reference. An average of the distances calculated for each interaction is compared in table 3 with interproton distances measured in several structures obtained from molecular dynamics calculations and energy minimizations (see next section). Errors in the experimental distances (table 3) were estimated from the scattering in the different values calculated, and from estimated errors associated with the methodology (77) and with the measurement of cross-peak volumes.

An inspection of the distances measured by NOE quantitation (table 3) shows that the ones corresponding to the Pro 2-Ala 3 region of the cyclic hexapeptide fit very well with those expected for a standard Type I β turn (figure 1), with the exception of the Pro 2 H ^{α} /Ala 3 NH interaction. As mentioned above, this is very likely due to the presence of a small amount of a Type II β -turn structure in fast equilibrium with the main conformation of the peptide (40). Assuming that this distance is 3.5 Å and 2.1 Å for standard Type I and Type II β turns, respectively, the population of Type II conformation can be estimated to be 7%. In the next sections we compare in detail the experimental data with results from molecular dynamics simulations and will return to this point.

Molecular Dynamics Simulations

Nuclear magnetic resonance is a powerful tool to determine the structure of a molecule, but the information obtained corresponds, in general, to an average over an ensemble of conformations visited by the molecule in the NMR time scale, conformations that can differ to a variable extent from one another. *In vacuo* molecular dynamics and energy minimizations provide complementary information that can help to fit the NMR data into a single family of conformations or can lead to several structures that may contribute to a greater or lesser degree to the experimental observations (40). Although the lack of solvent in these calculations is a serious limitation, they are very useful in exploring the conformational states accessible to

Table 3. Interproton distances (Å) in cyclo(Asn-Pro-Ala-Leu-Pro-Gly): measured by NMR and calculated in structures obtained by molecular dynamics (see text). The constraints imposed in one of the molecular dynamics simulations, the calculated energy in the molecular dynamics structures, and the distances between atoms implicated in hydrogen bonds in these structures, are indicated.

Distance	NMR (a)	Structure 1	Structure 2	Structure 3
N1 NH/G6 NH (b)	3.0 (2.8–3.3)	2.79	3.30	2.92
A3 NH/L4 NH (b)	2.4 (2.2–2.6)	2.46	2.52	2.48
N1 NH/NH ^{δ2}	3.2 (2.9–3.5)	2.86	4.43	4.89
G6 H ^α /N1 NH	2.8 (2.5–3.1) (c)	2.52	2.45	2.57
	3.1 (2.8–3.5) (d)	3.50	3.57	3.48
P2 H ^α /A3 NH	3.0 (2.8–3.3)	3.51	3.55	3.54
A3 H ^α /L4 NH (b)	3.1 (2.8–3.4)	3.59	3.32	3.57
P5 H ^α /G6 NH (b)	2.3 (2.1–2.5)	2.00	2.05	1.99
N1 NH/H ^α (b)	2.3 (2.1–2.6)	3.05	2.77	3.04
A3 NH/H ^α	2.7 (2.5–3.0)	3.06	3.06	3.05
L4 NH/H ^α (b)	2.8 (2.5–3.1)	3.07	2.99	3.04
G6 NH/H ^α	2.6 (2.4–2.9) (c)	2.78	2.76	2.73
	2.9 (2.6–3.2) (d)	3.04	3.05	3.06
N1 NH/H ^β	2.3 (2.1–2.6) (c)	3.19	2.46	2.45
	2.6 (2.3–2.9) (d)	3.92	2.33	2.80
N1 NH ^{δ2} /H ^β	2.6 (2.3–2.9) (c)	2.29	3.44	3.45
	3.0 (2.7–3.3) (d)	3.44	2.25	2.27
L4 NH/N1 H ^β	3.1 (2.8–3.4) (c)	4.95	2.63	2.41
	3.5 (3.1–3.9) (d)	4.52	4.34	4.12
A3 NH/P2 H ^δ (e)	2.8 (2.5–3.2)	3.24	3.30	3.28
A3 NH/P2 H ^β (f)	2.8 (2.5–3.1)	3.03	2.82	2.83
N1 H ^β /P2 H ^δ (e)	2.7 (2.4–3.0) (c)	2.31	4.36	4.34
	3.0 (2.7–3.3) (d)	2.96	4.32	4.33
N1 H ^α /P2 H ^δ (b, g)	2.1 (2.0–2.3)	2.25	2.30	2.24
N1 H ^α /P2 H ^δ (b, e)	2.2 (2.0–2.4)	3.24	2.78	2.84
L4 H ^α /P5 H ^δ (b)	2.2 (2.0–2.5) (d)	2.25	2.25	2.30
		2.53	2.56	2.51
N1 H ^α /H ^β	2.4 (2.2–2.6) (d)	2.46	3.06	3.10
		2.40	2.54	2.56
N1 NH/L4 O		2.03	2.38	2.01
LA NH/N1 O			2.21	2.13
A3 NH/N1 O ^δ (b)		2.31	2.15	2.19
LA NH/N1 O ^δ		2.12		
Energy (kcal/mol)		79.2	86.6	80.0

(a) Distances measured by cross-peak volume quantitation on one NOESY and two ROESY spectra at short mixing times (see text and methods for details). As some important NOEs involve equivalent methylene protons, limiting distance ranges have been calculated assuming that only one of them contributes to the NOE (c) or that both contribute to the same extent (d). In the molecular dynamics structures, the distances to both methylene protons are indicated. (b) Distances used as constraints in one of the molecular dynamics simulations as explained in methods. (c) Distance calculated assuming that only one of the two equivalent methylene protons contributes to the NOE. (d). Distance calculated assuming that both equivalent methylene protons contribute to the same extent to the NOE. (e) Upfield Pro 2 H^δ. (f) Upfield Pro 2 H^β. (g) Downfield Pro 2 H^δ.

the molecule, and interproton distance constraints can be imposed to satisfy better the NMR data while maintaining the energy requirements of the molecule. We have used the program DISCOVER (73) to carry out molecular dynamics simulations and energy minimizations of the cyclic hexapeptide. In combination with the interproton distance data, the information afforded by these calculations has enabled us to refine the model delineated for the molecule and have a better picture of its conformational behavior.

To carry out an initial search of the conformational space available to cyclo(Asn-Pro-Ala-Leu-Pro-Gly) two different structures of the peptide were built-up and minimized, and two 40 ps *in vacuo* molecular dynamics trajectories at 600 K were calculated. Instantaneous structures taken at 1 ps intervals from the trajectories were energy minimized and characterized by their backbone and Asn 1 side chain torsion angles, and by their hydrogen bonding patterns. The first starting structure contained Type I β turns around Pro 2-Ala 3 and Pro 5-Gly 6, and the Asn 1 side chain was located above the ring. In the trajectory started with this structure, a conversion of the Pro 5-Gly 6 turn from Type I to Type II was observed during the first 5 ps, and no further conversion back to a Type I conformation was observed during the remaining 35 ps. On the other side of the molecule, oscillations between Type I and Type II β -turn structures were observed during the beginning of the simulation, but the conformation was stabilized in a Type I turn after the Asn 1 amide group rotated and established favorable interactions with the Ala 3 NH. The most frequent of the minimized structures obtained from this simulation contains

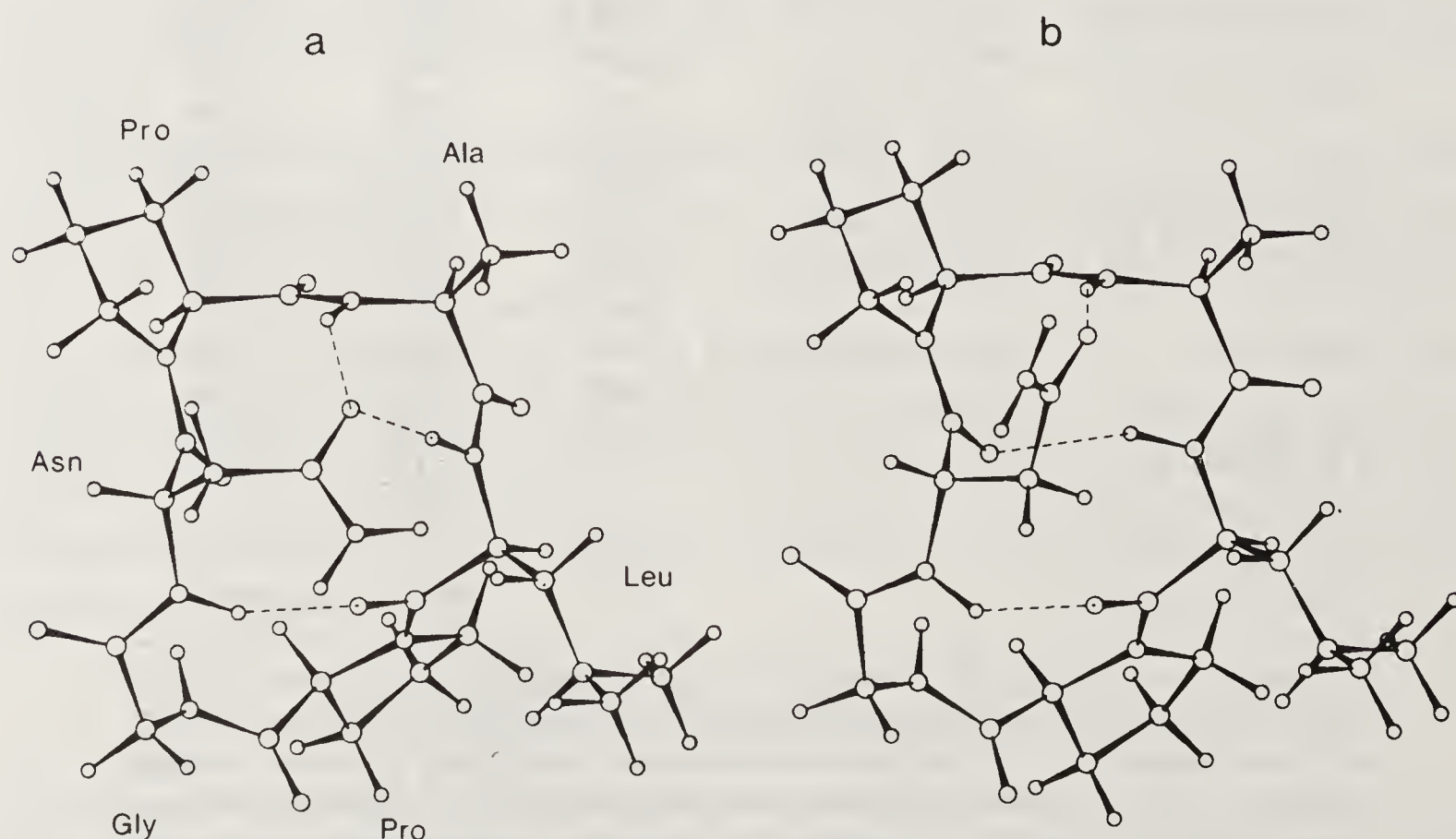


Figure 6. Two structures of cyclo(Asn-Pro-Ala-Leu-Pro-Gly) with different Asn side chain conformations, obtained by molecular dynamics and energy minimization. (a) Structure 1 (table 3); (ϕ, ψ) dihedral angles: ($-91.5, 171.6$), ($-56.7, -31.8$), ($-87.8, -56.3$), ($-98.1, 117.2$), ($-43.2, 118.1$), ($125.5, -66.4$); Asn (χ_1, χ_2): ($71.4, -102.6$). (b) Structure 2 (table 3); (ϕ, ψ) dihedral angles: ($-57.3, 144.8$), ($-47.5, -36.8$), ($-91.7, -9.9$), ($-148.1, 115.8$), ($-32.4, 105.7$), ($123.1, -76.2$); Asn(χ_1, χ_2): ($-166.8, 96.4$). Hydrogen bonds are indicated by dashed lines.

a hydrogen bonded Type II β turn with Pro 5–Gly 6 in the corner positions, and a Type I β turn around Pro 2–Ala 3 stabilized by an Asx turn. The Leu 4 NH participates with Ala 3 NH in a bifurcated hydrogen bond with the Asn 1 O $^{\delta}$, rather than being hydrogen bonded to Asn 1 O. This structure, the minimum energy conformation that we have found up to now, is represented in figure 6 and the interproton distances obtained from it are compared in table 3 (structure 1) with the experimental ones.

The second starting structure consisted of a Type I β turn around Pro 2–Ala 3 and a Type II β turn with Pro 5–Gly 6 in the corner positions, and the Asn 1 side chain was extended beside the ring. This overall conformation was maintained during the whole trajectory, although, lacking the Asn 1 O $^{\delta}$ /Ala 3 NH interaction, higher oscillations around the Pro 2–Ala 3 peptide bond were observed and conformations close to a Type II turn were sometimes briefly visited. The Asn 1 side chain showed high mobility and, although seldom, interactions between the NH $^{\delta 2}$ and all but the Asn 1 backbone oxygens of the peptide were observed. Hydrogen bonds with the Pro 2, Ala 3 and Gly 6 carbonyls were maintained in different minimized structures obtained from the simulation, whose energies were 5–9 kcal/mol higher than the energy of structure 1. A hydrogen bond between Gly 6 NH and Leu 4 O was frequently observed in the trajectory and in the minimized structures, suggesting that the conformational space sampled during this simulation could be representative of the structure existing in DMSO-d₆ solution.

To better characterize the lowest energy conformation found in the first simulation (structure 1) and the conformational space near this structure, a 40 ps molecular dynamics trajectory was calculated at 300 K. Instantaneous structures taken at 0.1 ps intervals along the simulation are represented in figure 7a in the form of Ramachandran maps for every residue of the peptide, and the most frequent hydrogen bonds observed in these structures are represented as a function of time in figure 8a. As can be observed in figure 7a, the molecule remained fundamentally in the same conformational family during the whole trajectory, including the Asn 1 side chain. Correspondingly, structure 1 and a very similar conformation were the only minima obtained after energy minimization of 40 structures visited during the simulation. The most frequent hydrogen bonds observed along the trajectory (figure 8a) are the ones contained in structure 1, but an interaction between Leu 4 NH and Asn 1 O was also often formed.

Comparing the interproton distances measured experimentally with the ones calculated for structure 1 (table 3), a general good fit is observed, but some points of discrepancy are observed. To try to obtain a better fit of the NMR distances, a 40 ps constrained molecular dynamics simulation at 900 K was carried out starting from structure 1. The constraints used in the simulation are also indicated in table 3, and included the Asn 1 O $^{\delta}$ /Ala 3 NH hydrogen bond. Several minimized structures were obtained along the trajectory, and compared. The structure with the lowest energy (structure 2, table 3 and figure 6) was the one that best satisfied the constraints and only the Asn 1 NH/H $^{\alpha}$ distance fell out of the range imposed. This structure contains hydrogen bonds between Asn 1 NH/Leu 4 O, Leu 4 NH/Asn 1 O and Ala 3 NH/Asn 1 O $^{\delta}$, and the main difference with respect to structure 1 is in the location of the Asn 1 side chain, which is visiting another rotameric state.

To relax the structure obtained from constrained molecular dynamics, an energy minimization was conducted without the constraints, and a 40 ps trajectory was

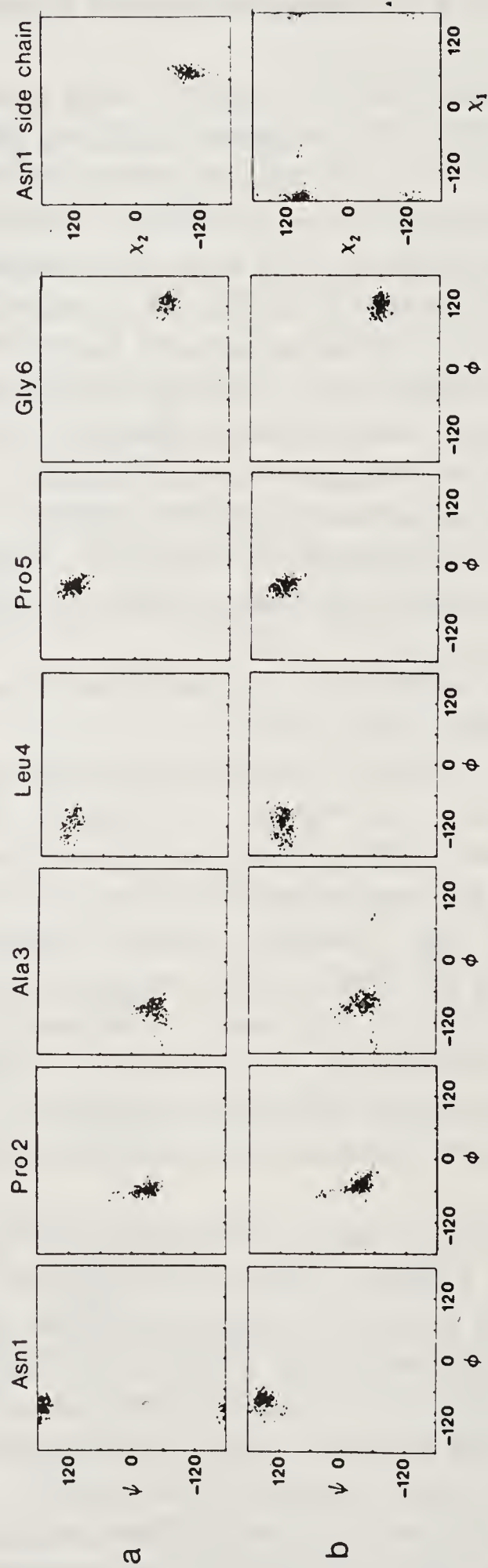


Figure 7. Ramachandran maps representing the conformations visited by cyclo(Asn-Pro-Ala-Leu-Pro-Gly) during two 40 ps molecular dynamics simulations (every dot corresponds to one instantaneous structure taken at 0.1 ps intervals along the trajectory). (a) Simulation started from structure 1 (table 3 and figure 6). (b) Simulation started after non-constrained minimization of structure 2 (table 3 and figure 6).

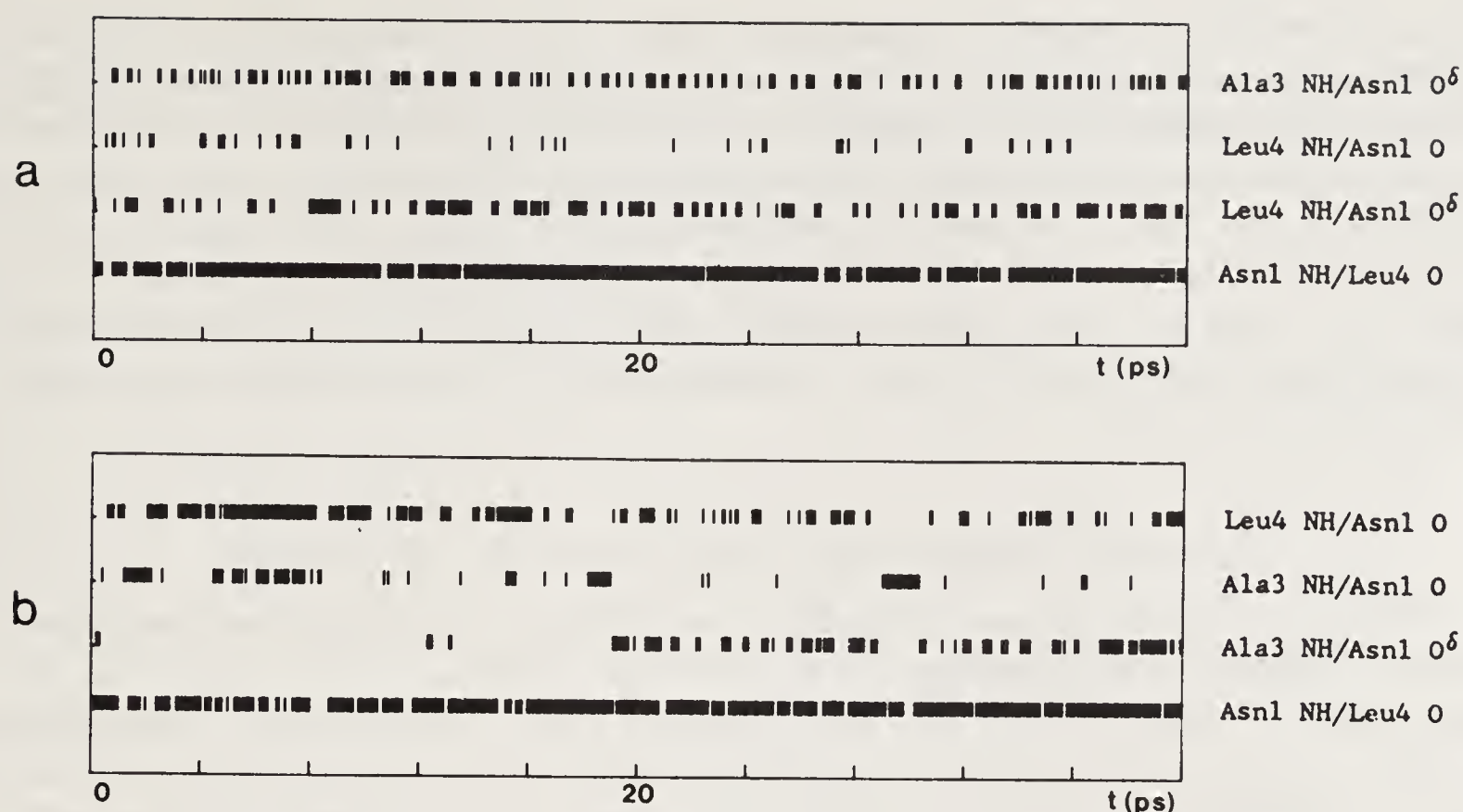


Figure 8. Summary of the most frequent hydrogen bond interactions formed during two 40 ps molecular dynamics simulations of cyclo(Asn-Pro-Ala-Leu-Pro-Gly). Every tick mark corresponds to a hydrogen bond observed in instantaneous structures taken at 0.1 ps intervals along the trajectory. (a) Simulation started from structure 1 (table 3 and figure 6). (b) Simulation started after non-constrained minimization of structure 2 (table 3 and figure 6).

calculated at 300 K starting from the minimized conformation. The conformational states visited during the simulation are summarized in figure 7b, and the most frequent hydrogen bonds observed in the trajectory are represented in figure 8b. A change in the rotameric state of the Asn 1 side chain occurred in the initial stages of the simulation, where the hydrogen bond between Ala 3 NH and Asn 1 O $^{\delta}$ was disrupted, but another conformational change brought the molecule back to the initial conformation and this hydrogen bond was maintained during the last 20 ps of the trajectory (figure 8b). A hydrogen bond between Leu 4 NH and Asn 1 O was frequently formed along the whole simulation, in particular when the interaction between the Asn 1 side chain and the Ala 3 NH was not present. The absence of this hydrogen bond also seems to favor an interaction between Ala 3 NH and Asn 1 O. The hydrogen bond closing the Pro 5-Gly 6 Type II β turn is almost permanently observed in all the simulations described above, showing the stability of this conformation. The most frequent structure obtained by energy minimization along the trajectory (structure 3, table 3) is very similar to the constrained conformation (structure 2), but lacks the hydrogen bond between Leu 4 NH and Asn 1 O. The energy of this structure is 0.8 kcal/mol higher than the one of structure 1.

A final 40 ps molecular dynamics simulation at 300 K was performed to assess the possibility that a Type II β -turn structure with Pro 2-Ala 3 in the corner positions could be contributing significantly to the NMR observations for cyclo(Asn-Pro-Ala-Leu-Pro-Gly) in CDCl₃ (14% DMSO-d₆). The starting structure was obtained from structure 1 by torsional forcing Pro 2 ψ to 120° (40). The Pro 2-Ala 3 region of the molecule remained in a Type II β turn conformation for the first 30 ps of the simulation and then returned to the conformational space of structure 1. Several energy minima

with a Type II conformation around Pro 2–Ala 3 were found, with 5–7 kcal/mol higher energy than structure 1. These conformations are, therefore, local energy minima that could be present in solution, in particular in the presence of a hydrogen bond acceptor that can favor (in relative terms) conformations with more exposed amide protons. This conclusion is supported by the increase in the Type II β -turn conformation around Pro 2–Ala 3 observed when the amount of DMSO- d_6 in the solvent is raised to 100%, and correlates with results described above for model dipeptides that show Type I β -turn conformations in chloroform solution and Type II turns in crystals (33).

Comparison between NMR and Molecular Dynamics Results

The interproton distances calculated for structures 1–3 are compared in table 3 with the distances measured by NOE quantitation. Structures 2 and 3 can be considered to belong to the same conformational family and differ from structure 1 mainly in the location of the Asn 1 side chain. As can be observed in figure 7, the peptide backbone is visiting essentially the same conformational space in both structural families, with just a small shift in the Asn 1 ϕ dihedral angle. Correspondingly, the largest differences in the interproton distances calculated from structure 1 and from structures 2 and 3 involve Asn 1 side chain protons. Comparing with the distances measured by NMR, we observe that structure 1 cannot account for the observed NOE between Leu 4 NH and the Asn 1 H^β protons. The Asn 1 side chain is also too far from the Leu 4 side chain to explain the NOEs observed between them (figure 4), which have not been quantitated. These NOEs involving the Asn 1 H^β protons are better fit by structures 2 and 3, while the NOEs between Asn 1 H^β and Pro 2 H^δ , and between Asn 1 NH and Asn 1 $NH^{\delta 2}$ can only be explained by structure 1. The observation of all these NOEs, and the magnitudes of the interproton distances deduced from them, clearly point to a coexistence of both Asn 1 side chain conformations in fast equilibrium on the NMR time scale. The measured Asn 1 NH/ H^β and Asn 1 H^α / H^β distances are also suggestive of the presence of both conformations in the experimental conditions of this study.

In general, structures 1–3 fit quite well the measured interproton distances involving backbone protons. In particular, the Ala 3 NH/Leu 4 NH and Gly 6 NH/Asn 1 NH distances are reproduced very well. As mentioned above, some experimental data can be better explained on the assumption that a small amount of conformations containing a Type II β turn around Pro 2–Ala 3 also exists (e.g. Ala 3 NH/Pro 2 H^α and Ala 3 NH/Ala 3 H^α). Some small discrepancies are found in the Pro 5–Gly 6 region of the molecule, but the overall conformation corresponds to the NMR observations. The molecular dynamics simulations indicate that the Type II β -turn structure is very stable, but the contribution of a small percentage of Type I conformations around Pro 5–Gly 6 to the observed NOE pattern cannot be eliminated. The most significant difference between the calculated structures and the NMR data is in the Asn 1 ϕ dihedral angle. The small $^3J_{HN\alpha}$ coupling constant (table 1) and the short Asn 1 NH/ H^α distance measured point to smaller values (ca. -50°) than the ones calculated in structures 1–3 for this torsion angle. In fact, most of the energy expended to constrain structure 2 is used to constrain the Asn 1 NH/ H^α interproton distance. Finally, it is noteworthy that the distances between the Asn 1 H^α and Pro 2 H^δ protons, and between the Leu 4 H^α and Pro 5 H^δ protons, are longer, in general, in

the calculated structures than in the experimental values. Computer graphics show that the distance between the H^α proton of a residue preceding proline and the Pro H^δ protons can only be so short ($\sim 2.2 \text{ \AA}$) for the H^δ at a time. These observations suggest that some spin diffusion, or some sequential transfer via dipolar and scalar coupling, could be occurring, favored by the strong interaction between the Pro H^δ geminal protons. The observation of a small NOE between the Asn 1 $NH^{\delta 1}$ and H^β protons even for 0.1 s mixing time supports this conclusion.

The comparison of the NMR and molecular dynamics results obtained for cyclo(Asn-Pro-Ala-Leu-Pro-Gly) leads to the conclusion that the expected structure containing a Type I β turn around Pro 2-Ala 3 and a Type II β turn around Pro 5-Gly 6 is the main conformation of the peptide in $CDCl_3$ (14% DMSO- d_6) solution. The Asn 1 side chain can adopt at least two different conformations with a hydrogen bond with Ala 3 NH and can also interact with Leu 4 NH. Structures 1-3 can only be taken as close representations of the actual conformations of the cyclic hexapeptide in solution, as all the molecular dynamics simulations have been carried out *in vacuo*, but they can account for most of the experimental observations. The inclusion of NMR constraints in the calculations allowed the refinement of the peptide backbone conformation (structure 2), but could not by itself explain all the measured interproton distances. Without consideration of several conformational possibilities and a careful choice of the distances that can be used as constraints, completely misleading results could have been obtained. The results presented here for the cyclic hexapeptide, and our previous results for cyclic pentapeptides (40), indicate that the contribution of more than one conformation to the experimental observations should be considered in general, even for molecules expected to be fairly rigid.

The hydrogen bonding patterns observed in structures 1-3 (table 3) explain quite well the temperature dependence data obtained by NMR (table 1). However, while it seems clear that the Asn 1 NH is hydrogen bonded to Leu 4 O, and Ala 3 NH to Asn 1 O^δ , a question remains about the interactions involving the Leu 4 NH. The NMR data suggest that this amide proton is hydrogen bonded to Asn 1 O, and inclusion of NMR constraints to molecular dynamics calculations (structure 2) leads to the formation of this interaction. This hydrogen bond does not appear in non-constrained minimized conformations (structures 1 and 3), but is frequently formed in molecular dynamics simulations (figure 8), and is replaced by a Leu 4 NH/Asn 1 O^δ in the minimum energy conformation (structure 1). It seems likely that the Leu 4 NH/Asn 1 O hydrogen bond is mostly present in solution and that favorable interactions with the Asn 1 side chain oxygen occur, leaving this amide proton highly hindered from the solvent, as is observed by NMR (table 1). Hence, the idea is reinforced that a dynamic view of the structural models delineated for a molecule will be always more appropriate to understand its behaviour than the consideration of only static conformations.

Conclusion

Our cyclic hexapeptide model reproduces quite well the conformation of the Asn 148-Pro-Ala-Leu 151 sequence of dihydrofolate reductase (*L. casei*) in the protein crystals, including the Type I β turn and the Asn side chain/Ala NH interaction.

However, it is noteworthy that the conformation of the Asn side chain in the protein, with both dihedral angles close to 180° , does not fit with the conformations found in our study for the cyclic hexapeptide. This indicates that several conformations of the Asn side chain can accommodate an interaction between the Asn O^δ in the i position of a Type I β turn and the NH of residue $i + 2$ in the turn. Although only one of these conformations is maintained in the crystal, other conformations that maintain this hydrogen bond can exist in solution and can confer additional stability to the β turn. Our results suggest that a favorable interaction between the Asn (i) side chain and the NH of residue $i + 3$ can also stabilize the turn. In agreement with this possibility is the fact that interactions between the O^δ of asparagine residues (n) and the NH of residue $n + 3$, and bifurcated hydrogen bonds from Asn (i) side chains to the amide protons of residues $i + 2$ and $i + 3$ in Type I β turns (such as in structure 1, figure 6), have both been observed in protein crystals (7, 19). With no doubt, all these interactions between the Asn side chain and the amide protons of the next residues in the sequence will not only stabilize turns in folded protein structures, but also increase the population of turn conformations in the flexible chain of an unfolded protein. Hence, the formation of turns with Asn in the i position can play an active role in the folding of a protein.

Acknowledgements

This work was supported by grants from the NIH (GM 27616) and from the Robert A. Welch Foundation. JR was supported by a grant from the Ministerio de Educacion y Ciencia, Spain.

References

1. Richardson, J. S. (1981) *Adv. Protein Chem.* **34**, 167–339.
2. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22**, 2577–2637.
3. Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985) *Adv. Protein Chem.* **37**, 1–109.
4. Geddes, A. J., Parker, K. D., Atkind, E. D. T. & Beighton, E. (1968) *J. Mol. Biol.* **32**, 343–358.
5. Venkatachalam, C. M. (1968) *Biopolymers* **6**, 1425–1436.
6. Chou, P. Y. & Fasman, G. D. (1977) *J. Mol. Biol.* **115**, 135–175.
7. Wilmot, C. M. & Thornton, J. M. (1988) *J. Mol. Biol.* **203**, 221–232.
8. Wilmot, C. M. & Thornton, J. M. (1990) *Protein Eng.*, **3**, 479–493.
9. Nemethy, G. & Printz, M. P. (1972) *Macromolecules* **5**, 755–758.
10. Mathews, B. W. (1972) *Macromolecules* **5**, 818–819.
11. Kuntz, I. D. (1972) *J. Am. Chem. Soc.* **94**, 4009–4012.
12. Lewis, P. N., Momany, F. A. & Scheraga, H. A. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 2293–2297.
13. Rose, G. D., Winters, R. H. & Wetlaufer, D. B. (1976) *FEBS Lett.* **63**, 10–16.
14. Lesk, A. M. & Rose, G. D. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4304–4308.
15. Ptitsyn, O. B. (1981) *FEBS Lett.* **131**, 197–202.
16. King, J. (1986) *Biotechnol.* **4**, 297–303.
17. Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A. & Wright, P. E. (1988) *J. Mol. Biol.* **201**, 161–200.
18. Wright, P. E., Dyson, H. J. & Lerner, R. A. (1988) *Biochemistry* **27**, 7167–7175.
19. Baker, E. N. & Hubbard, R. E. (1984) *Prog. Biophys. Mol. Biol.* **179**, 97–179.
20. Maigret, B. & Pullman, B. (1974) *Theor. Chim. Acta* **35**, 113–128.
21. Peters, D. & Peters, J. (1980) *J. Mol. Struct.* **62**, 229–247.
22. Chuman, H., Momany, F. A. & Schafer, L. (1984) *Int. J. Pept. Protein Res.* **24**, 233–248.

23. Karle, I. L. (1981) in *The Peptides* eds. Gross, E. & Meienhofer, J. (Academic Press, New York), Vol. 4, pp. 1–54.
24. Kessler, H. (1982) *Angew. Chem., Int. Ed. Engl.* **21**, 512–523.
25. Woody, R. W. (1974) in *Peptides, Polypeptides and Proteins*, eds. Blout, E. R., Bovey, F. A., Goodman, M. & Lotan, N. (Wiley, New York), pp. 338–350.
26. Krimm, S. (1983) *Biopolymers* **22**, 217–225.
27. Seaton, B. (1986) *Spectrochim. Acta* **42A**, 227–232.
28. Scheraga, H. A. (1981) *Biopolymers* **20**, 1877–1899.
29. Hagler, A. T. (1985) in *The Peptides*, eds. Hruby, V. & Meienhofer, J. (Academic Press, New York), Vol. 7, pp. 213–299.
30. Aubry, A. & Marraud, M. (1983) *Biopolymers* **22**, 341–345.
31. Boussard, G., Marraud, M. & Aubry, A. (1979) *Biopolymers* **18**, 1297–1331.
32. Boussard, G. & Marraud, M. (1985) *J. Am. Chem. Soc.* **107**, 1825–1827.
33. Aubry, A., Cung, M. T. & Marraud, M. (1985) *J. Am. Chem. Soc.* **107**, 7640–7647.
34. Aubry, A., Ghermani, N. & Marraud, M. (1984) *Int. J. Pept. Protein Res.* **23**, 113–122.
35. Marraud, M. & Aubry, A. (1984) *Int. J. Pept. Protein Res.* **23**, 123–133.
36. Mcharfi, M., Aubry, A., Boussard, G. & Marraud, M. (1986) *Eur. Biophys. J.* **14**, 43–51.
37. Aubry, A., Abbadi, A., Boussard, G. & Marraud, M. (1987) *New J. Chem.* **11**, 739–744.
38. Pichon-Pesme, V., Aubry, A., Abbadi, A., Boussard, G. & Marraud, M. (1988) *Int. J. Pept. Protein Res.* **32**, 175–182.
39. Abbadi, A., Boussard, G. & Marraud, M. (1986) *Int. J. Biol. Macromol.* **8**, 252–255.
40. Stradley, S. J., Rizo, J., Bruch, M. D., Stroup, A. N. & Gierasch, L. M. (1990) *Biopolymers* **29**, 263–287.
41. Narasinga Rao, B. N., Kumar, A., Balaram, H., Ravi, A. & Balaram, P. (1983) *J. Am. Chem. Soc.* **105**, 7423–7428.
42. Kishore, R., Ragothama, S. & Balaram, P. (1988) *Biochemistry* **27**, 2462–2471.
43. Garcia-Echeverria, C., Albericio, A., Pons, M., Barany, G. & Giralt, E. (1989) *Tetrahedron Lett.* **30**, 2441–2444.
44. Venkatachalapathi, Y. V., Prasad, B. V. V. & Balaram, P. (1982) *Biochemistry* **21**, 5502–5509.
45. Deslauriers, R., Evans, D. J., Leach, S. J., Meinwald, Y. C., Minasian, E., Nemethy, G., Rae, I. D., Scheraga, H. A., Somorjai, R. L., Stimson, E. R., Van Nispen, J. W. & Woody, R. W. (1981) *Macromolecules* **14**, 985–996.
46. Bandekar, J., Evans, D. J., Krimm, S., Leach, S. J., McQuie, J. R., Minasian, E., Nemethy, G., Pottle, M. S., Scheraga, H. A., Stimson, E. R. & Woody, R. W. (1982) *Int. J. Pept. Protein Res.* **19**, 187–205.
47. Mizuno, H., Lee, S., Nakamura, H., Kodera, Y., Kato, T., Go N. & Izumiya, N. (1986) *Biophys. Chem.* **25**, 73–90.
48. Hollosi, M., Kover, K. E., Holly, S., Radics, L. & Fasman, G. D. (1987) *Biopolymers* **26**, 1555–1572.
49. Bruch, M. D., Noggle, J. H. & Gierasch, L. M. (1985) *J. Am. Chem. Soc.* **107**, 1400–1407.
50. Gierasch, L. M., Deber, C. M., Madison, V., Niu, C.-H. & Blout, E. R. (1981) *Biochemistry* **20**, 4730–4738.
51. Kitson, D. H. & Hagler, A. T. (1988) *Biochemistry* **27**, 5246–5257.
52. Senn, H., Loosli, H.-R., Sanner, M. & Braun, W. (1990) *Biopolymers* **29**, 1387–1400.
53. Karten, M. & Rivier, J. (1986) *Endocr. Rev.* **7**, 44–66.
54. Goodman, M., Rone, R., Manesis, N., Hassan, M. & Mammi, N. (1987) *Biopolymers* **26**, S25–S32.
55. Schulze-Gahmen, U., Prinz, H., Glatter, U. & Beyreuther, K. (1985) *EMBO J.* **4**, 1731–1737.
56. Thomas, Jr., G. J., Prescott, B. & Urry, D. W. (1987) *Biopolymers* **26**, 921–934.
57. Jelicks, L. A., Naider, F. R., Shenbagamurthi, P., Becker, J. M. & Broido, M. S. (1988) *Biopolymers* **27**, 431–449.
58. Stroup, A. N. & Gierasch, L. M. (1990) *Biochemistry*, **29**, 9765–9771.
59. Yu, M.-H. & King, J. (1988) *J. Biol. Chem.* **263**, 1424–1431.
60. Rees, D., Lewis, M. & Lipscomb, W. (1983) *J. Mol. Biol.* **168**, 367–387.
61. Richardson, J. S. & Richardson, D. C. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation*, ed. Fasman, G. D. (Plenum Press, New York), pp. 1–98.

62. Pease, L. G. & Watson, C. (1978) *J. Am. Chem. Soc.* **100**, 1279–1286.
63. Davis, D. G. & Bax, A. (1985) *J. Am. Chem. Soc.* **107**, 2821–2823.
64. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979) *J. Chem. Phys.* **71**, 4546–4553.
65. Macura, S., Huang, Y., Suter, D. Ernst, R. R. (1981) *J. Magn. Reson.* **43**, 259–281.
66. Kumar, A., Wagner, G., Ernst, R. R. & Wuethrich, K. (1981) *J. Am. Chem. Soc.* **103**, 3654–3658.
67. Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D. & Jeanloz, R. W. (1984) *J. Am. Chem. Soc.* **106**, 811–813.
68. Kessler, H., Griesinger, C., Kerssebaum, R., Wagner, K. & Ernst, R. R. (1987) *J. Am. Chem. Soc.* **109**, 607–609.
69. States, D. J., Haberkorn, R. A. & Ruben, D. J. (1982) *J. Magn. Reson.* **48**, 286–292.
70. Bax, A. & Davis, D. G. (1985) *J. Magn. Reson.* **63**, 207–213.
71. FTNMR program, available from Hare Research Inc., Woodinville, WA.
72. Otting, G., Widmer, H., Wagner, G. & Wuethrich, K. (1986) *J. Magn. Reson.* **66**, 187–193.
73. DISCOVER, MOLEDT and INSIGHTII are commercially available from Biosym Technologies, 9605 Scranton Rd., San Diego, CA 92121.
74. Dorman, D. E. & Bovey, F. A. (1973) *J. Org. Chem.* **38**, 2379–2383.
75. Wuethrich, K. (1986) *NMR of Proteins and Nucleic Acids*, (Wiley, New York).
76. Gierasch, L. M., Karle, I. L., Rockwell, A. L. & Yenai, K. (1985) *J. Am. Chem. Soc.* **107**, 3321–3327.
77. Clore, G. M. & Gronenborn, A. M. (1985) *J. Magn. Reson.* **61**, 158–164.

The 3_{10} - and α -Helical Conformations in Peptides

Ettore Benedetti¹, Benedetto Di Blasio¹, Vincenzo Pavone¹,
Carlo Pedone¹, Antonello Santini¹, Marco Crisma²
and Claudio Toniolo²

¹Department of Chemistry, University of Naples,
80134 Napoli, Italy

²Biopolymer Research Centre, C.N.R., Department of Organic Chemistry,
University of Padova, 35131 Padova, Italy

Introduction

Among the peptide-ordered secondary structures occurring in the crystal state, helices represent a rather common observation. Structural details of these helices (1–5), with particular emphasis on deviations from the ‘ideal’ values that geometrical and conformational parameters might assume, are of relevance for our understanding of the forces responsible for their stabilization.

Most helices can be visualized as a succession of intramolecularly H-bonded cyclic structures of various types, such as the $3 \rightarrow 1$, $4 \rightarrow 1$, $5 \rightarrow 1$ or $6 \rightarrow 1$ structures, also called C_7 , C_{10} , C_{13} and C_{15} forms. In each of these cases the number of residues per turn, the pitch, the backbone torsion angles (φ and ψ), and the number of atoms in the cyclic structure stabilized by the intramolecular H-bond are different.

Recently, a considerable number of crystal structures of peptides with at least four aminoacid residues in the chain and folded in a helical conformation has appeared in the literature. However, no systematic analysis has been made so far to determine the parameters characterizing these helices.

We describe here the results of a general survey of all helices found in the X-ray crystal structure determination of peptides, the atomic coordinates of which are available up to June 1990. Our analysis, however, will focus on only two types of helical structures, namely the 3_{10} -helix and the α -helix, since in the sample of crystal structures examined they represent almost the totality of observed helical segments. Comparison will also be made with the results recently derived in a similar analysis by Barlow and Thornton (6) on the helix geometry observed in proteins.

Methods

Results from single crystal X-ray determinations only have been taken into account (those from fiber and powder work have not been considered). For inclusion in this

study the structures had to satisfy the following criteria:

- (i) Three-dimensional diffraction data from single crystal with no lattice disorder.
- (ii) Refined positional parameters for all non-hydrogen atoms of the peptide main chain.
- (iii) Absence of positional disorder in the main chain atoms of the peptide.
- (iv) Reasonable degree of refinement (conventional R value less than 0.10 with e.s.d.'s on bond lengths and bond angles no greater than 0.04 Å and 4.0°).
- (v) Presence of four residues in the chain with at least three of them showing a helical succession of conformational parameters (φ, ψ), either of the 3_{10} - or α -helical type.

The total number of independent molecules analyzed is 54, 12 of which contain exclusively the helicogenic Aib (α -aminoisobutyric acid) or Aib-like (Ac_5c , Ac_6c , etc.) residues (7). The overwhelming majority of the peptides are blocked at both N- and C-termini. The sample includes a total of 407 residues, 258 of which have been considered 'helical'. This number is still too small for a reliable statistical analysis subdivided according to type of residue. However, it is worth noting that 199 residues of the sample (corresponding to 49%) are Aib or Aib-like residues, 149 of which (corresponding to 36%) are part of a helical segment. A list of all crystal structures surveyed with the pertinent reference is available from E.B. upon request.

In table 1 statistical data are listed for subclasses of peptides divided according to the number of their residues. The total number of helical residues and their subdivision in α - and 3_{10} -helical residues together with the Aib (or Aib-like) content are reported. A total of 32 and 22 segments of the 3_{10} - and α -helical types, respectively, were analyzed.

The nomenclature and conventions used follow the recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature (8). Only right handed helices were considered in the analysis. Consequently, in crystal structures of peptides without chiral residues (as in oligopeptides composed of Aib or Aib-like residues only), in which right- and left-handed helices coexist in the unit cell of the crystal, only the right-handed helix was considered. In crystals of chiral peptides in which helices of both right- and left-handedness are found, both observed molecules were considered in the

Table 1. Statistical data for the helical segments analyzed in this survey

Peptide length	No. of molecules analyzed	No. of residues	No. of helical residues	% of helical residues	No. of α -helical residues	No. of 3_{10} -helical residues	% of Aib or Aib-like residues
Tetra	12	48	36	75	0	36	73
Penta	9	45	33	73	0	33	58
Hexa	3	18	14	78	0	14	83
Hepta	7	49	37	75	8	29	37
Octa	7	56	37	66	17	20	54
Nona	2	18	9	50	9	0	33
Deca	12	120	76	63	67	9	41
Undeca	1	11	7	64	7	0	36
Dodeca	1	12	7	58	7	0	50
Tetradeca	1	13	7	50	7	0	28
Hexadeca	1	16	5	31	5	0	31

analysis, but the signs of the (φ, ψ) torsion angles of the left-handed molecule were inverted. An inversion of sign of (φ, ψ) angles was also made in the case of chiral peptides forming exclusively a left-handed helix.

Results and Discussion

An inspection of the data given in table 1 leads to the following conclusions:

- (i) The minimal peptide main-chain length required for the formation of an α -helix corresponds to seven residues.
- (ii) Aib content of less than 50% favours the onset of an α -helix, but a 3₁₀-helix may still be observed at these Aib percentages.
- (iii) In nona- and longer peptides the α -helix is definitely preferred with respect to the 3₁₀-helix. However, at either end of the α -helical stretch one or two 3₁₀-helical residues are commonly observed.
- (iv) For peptides with seven residues or less, the observed helix comprises at least 75% of the residues. For octa or nonapeptides, on the average, half of the residues are in helical conformation. Actually, in this second group the number of helical residues is definitely greater, often because almost the entire molecule is helical. However, these helices are characterized by several irregularities represented by interruption of the main stretch of the helix and continuation of it at either end with one or two residues with different types of helical and H-bonding parameters.
- (v) The average number of residues in a helix shown by deca- or longer peptides is seven, corresponding to a helix of approximately two turns (with an end-to-end distance generally between 11 and 14 Å).

In table 2 the averaged conformational parameters (φ and ψ) for the two types of helices (α - or 3₁₀-helix) observed in peptides are listed together with the values of various “ideal” structures (2, 4, 9) and those recently published for proteins (6). The conformation of the observed helices is significantly different from the “ideal”

Table 2. A comparison between “standard” conformational parameters and those found in proteins and peptides

	φ	Ψ
A. α -Helices		
Pauling (2)	− 48	− 57
Perutz (4)	− 67	− 44
Arnott & Wonacott (9)	− 57	− 47
Proteins (6)	− 62	− 41
Peptides	− 63	− 42
B. 3 ₁₀ -Helices		
Pauling (2)	− 74	− 4
Perutz (4)	− 49	− 26
Proteins (6)	− 71	− 18
Peptides	− 57	− 30

structures. The differences observed in the main-chain conformational angles result in a greater outward tilt of the carbonyl groups from the helix axis (figure 1).

The mean (φ, ψ) conformational angles for right-handed α -helices found in peptides and proteins are very close, being ($-63^\circ, -42^\circ$) and ($-62^\circ, -41^\circ$), respectively. On the contrary, for 3_{10} -helices the conformational angles observed for peptides and

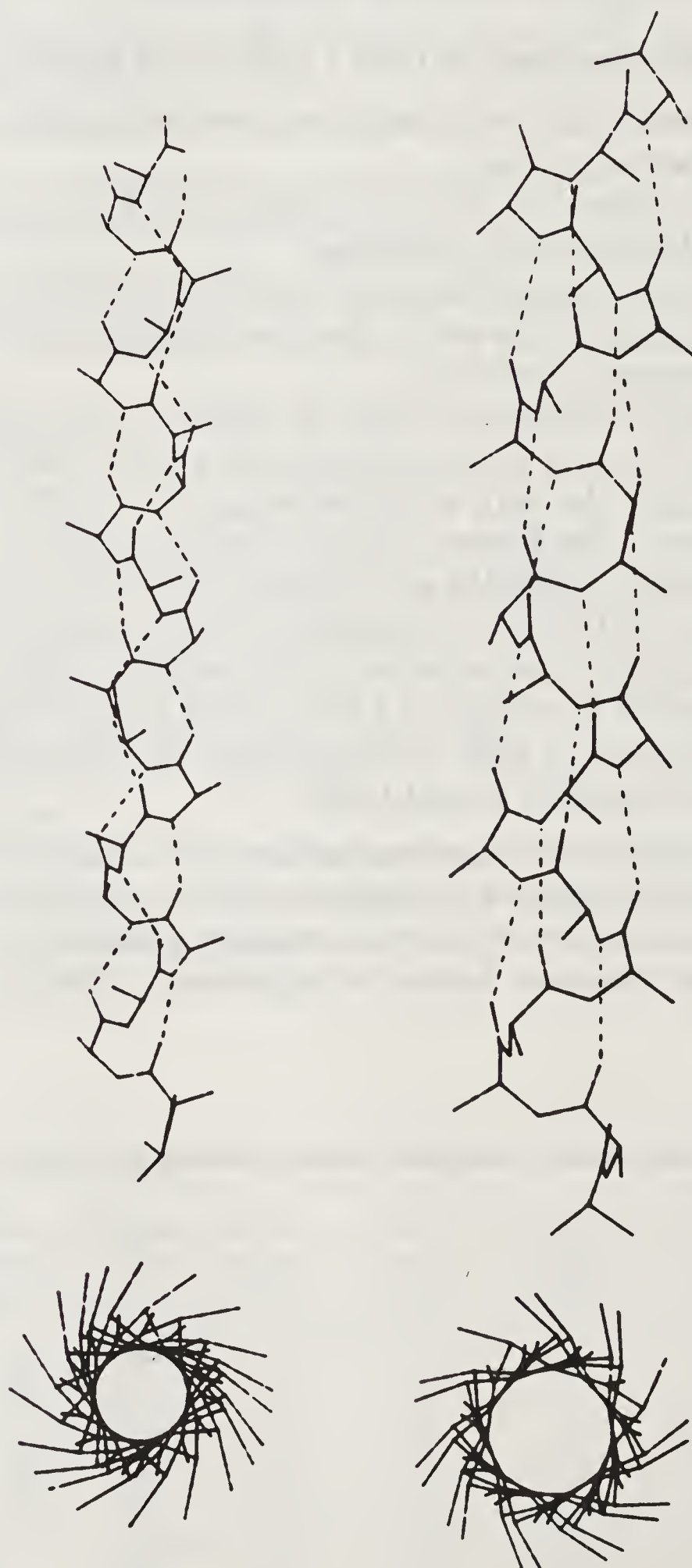


Figure 1. The 3_{10} -helix and the α -helix (left and right, respectively) as viewed perpendicularly (above) and along (below) the helix axis. The drawings have been calculated using standard geometries and the helical parameters obtained in this study.

proteins differ substantially, being $(-57^\circ, -30^\circ)$ and $(-71^\circ, -18^\circ)$, respectively. It is worth noting that in the case of peptides the 3_{10} -helices are longer than those observed in proteins (without considering tetrapeptides, the average length of 3_{10} -helices in peptides is of 4.9 residues). In fact in proteins only 3.4% of the residues are involved in 3_{10} -helices and the majority of the 3_{10} -helices have a mean length of 3.3 residues. In addition the peptide 3_{10} -helices are much more regular. In fact, the spread of the (φ, ψ) values is in the range $(-40^\circ, -80^\circ)$ and $(-5^\circ, -50^\circ)$, respectively, with 85% of both φ and ψ values being within 15° away from the calculated average value.

In table 3 the helical parameters θ , angle of twist per residue, h , height per residue, n , number of residues per turn and p , pitch of the helix, are reported for each subclass of peptides analyzed, together with the calculated weighted average parameters and those of "ideal" helices (2, 4, 9) and proteins (6).

The differences observed between the averaged helical parameters of peptides and proteins are small in both α - and 3_{10} -helices. For α - and 3_{10} -helices the peptide analysis shows values of both n and p slightly higher than those of proteins, but all within the calculated standard deviations. However, in the case of 3_{10} -helices we believe that the differences observed in the helical parameters are more significant. This has to be ascribed to the fact that our statistical sample is much more representative since in proteins only very small segments of 3_{10} -helices have been detected, while in peptides those segments are significantly longer (see above). Furthermore, the protein X-ray diffraction analysis has limitations because of the overall lower atomic resolution with respect to that obtained in crystal structures of peptides.

In summary, this crystallographic study of peptides has allowed us to characterize the 3_{10} -helix in great detail (at atomic resolution). For this helix the observed average n

Table 3. Average helical parameters for each subclass of peptides, overall average values weighted for the number of observations, and comparison with literature helical parameters

	3_{10} -Helix				α -Helix			
	θ , angle of twist per res. (in deg.)	h , height per res. (Å)	n , no. of residues per turn	p , pitch (Å)	θ , angle of twist per res. (in deg.)	h , height per res. (Å)	n , no. of residues per turn	p , pitch (Å)
Tetra	110.4	1.93	3.26	6.31	—	—	—	—
Penta	110.5	1.91	3.26	6.24	—	—	—	—
Hexa	109.8	1.98	3.27	6.46	—	—	—	—
Hepta	110.4	1.95	3.36	6.36	96.2	1.55	3.74	5.82
Octa	114.7	1.96	3.14	6.14	99.4	1.57	3.62	5.67
Nona	—	—	—	—	100.0	1.58	3.59	5.69
Deca	112.1	1.96	3.21	6.29	99.0	1.57	3.63	5.70
Higher	—	—	—	—	99.2	1.54	3.62	5.57
Weighted average*	111.1 (5.0)	1.94 (0.10)	3.24 (0.08)	6.29 (0.40)	99.0 (3.0)	1.56 (0.04)	3.63 (0.06)	5.67 (0.35)
Proteins (6)	112.5	1.81	3.2	5.8	101.7	1.52	3.54	5.4
Pauling (2)	120.0	2.00	3.0	6.0	98.6	1.51	3.65	5.5
Perutz (4)	112.5	1.93	3.0	5.8	98.1	1.42	3.67	5.2
Arnott & Wonacott (9)	—	—	—	—	100.3	1.53	3.59	5.5

*Estimated standard deviations (e.s.d.) for θ , h , n , and p values of 3_{10} - and α -helices are given in parentheses.

value (number of residues per turn) is 3.24. Consequently, the experimentally observed 3_{10} -helix has to be considered intermediate between the 'ideal' 3.0_{10} -helix ($n = 3.0$) and an 'ideal' α -helix ($n = 3.6$). As a result of this, the 3_{10} -helix shows an improved staggering of side-chains C^β atoms, as illustrated in figure 1.

This paper is dedicated to Prof. G. N. Ramachandran.

References

1. Bragg, W. L., Kendrew, J. C. and Perutz, M. F. (1950) *Proc. R. Soc. London* **A203**, 321–357.
2. Pauling, L., Corey, R. B. and Branson, H. R. (1951) *Proc. Natl. Acad. Sci. U.S.A.* **37**, 205–211.
3. Donohue, J. (1953) *Proc. Natl. Acad. Sci. U.S.A.* **39**, 470–478.
4. Perutz, M. F. (1951) *Nature (London)* **167**, 1053–1054.
5. Ramachandran, G. N., Venkatachalam, C. M. and Krimm, S. (1966) *Biophys. J.* **6**, 849–872.
6. Barlow, D. J. and Thornton, J. M. (1988) *J. Mol. Biol.* **201**, 601–619.
7. Toniolo, C. and Benedetti, E. (1988) *ISI Atlas of Science: Biochemistry* **1**, 225–230.
8. IUPAC-IUB Commission on Biochemical Nomenclature (1970) *Biochemistry* **9**, 3471–3479.
9. Arnott, S. and Wonacott, A. J. (1966) *J. Mol. Biol.* **21**, 371–383.

Helical Transitions in Peptides

John D. Clark,
Edward E. Hodgkin^a and Garland R. Marshall*

Center for Molecular Design, Washington University, St. Louis, MO 63130, USA

Introduction

While the α -helix is a recognized structural element in proteins, the role of the 3_{10} -helix is less well defined. A recent study of helix geometry in 57 globular proteins (1) indicated that the 3_{10} -helix motif is fairly common, but seldom in sequences of more than three or four residues. The definition of a helix of less than about five residues is somewhat questionable; however, the frequent appearance of the i to $i - 3$ hydrogen-bonding scheme with α -monoalkyl amino acids is noteworthy. Barlow and Thornton (1) also note that a helix generated with the average ϕ and ψ angles for 3_{10} -helical sections of globular proteins is essentially a ' $3\cdot2_{10}$ '-helix (i.e., a helix with 3.2 residues/turn and the ten-membered i to $i - 3$ hydrogen-bonding scheme) rather than the classically described helix with 3.0 residues/turn. A ' $3\cdot6_{10}$ '-helix (i.e., 3.6 residues/turn and the ten-membered i to $i - 3$ hydrogen-bonding scheme) has been described based on theoretical considerations (2). Recently, the crystal structure of the hirudin-thrombin complex has been solved and shows that the bound conformation of the C-terminal pentapeptide of hirudin exists in the 3_{10} -helix (3). Under the appropriate experimental conditions, this helix is also observed in the solution structure of hirudin itself[†], and it has been suggested (3) that the binding may be initiated by binding of this helix.

The presence of α,α -dialkyl amino acids, such as α -methylalanine (MeA[‡], aminoisobutyric acid, Aib), in microbial natural products, such as the peptaibol antibiotics, argues strongly for a special role related to function. One aspect is the conformational restrictions imposed by these amino acids (4–9). Despite the variety of

*For correspondence and reprint requests.

^aPresent address: British Biotechnology, Ltd., Oxford OX4 5LY, England.

Abbreviations: MeA, α -methylalanine or aminoisobutyric acid; NMR, nuclear magnetic resonance.

[†]Expanding Frontiers in Polypeptide and Protein Structural Research, International Biophysics Congress Satellite Symposium, Whistler, British Columbia, Canada, July 23–27, 1990, Abstract.

[‡]We prefer the use of 'MeA' to 'Aib', because of the increasing use of α -methyl derivatives of normal amino acids (see Marshall, G. R., Clark, J. D., Dunbar, J. B., Jr., Smith, G. D., Zabrocki, J., Redlinski, A. S., Leplawy, M. T. (1988) *Int. J. Pept. Protein Res.* 32, 544–555).

conformations theoretically available to α,α -dialkyl amino acids, the impact of multiple substitutions of these amino acids on the overall conformation of a peptide is a dramatic forcing of an α - or 3_{10} -helical conformation in most cases. The crystal structure of alamethicin (10), which contains eight MeA residues out of twenty, is predominately α -helical, with NMR data (11) supporting a similar solution conformation in methanol. According to Marshall *et al.* (12), those factors which govern helical preference include the inherent relative stability of the α -helix compared with the 3_{10} -helix, the extra hydrogen bond seen with 3_{10} -helices, and the enhanced electrostatic dipolar interaction of the 3_{10} -helix when packed in a crystalline lattice. The balance of these forces, when combined with the steric requirements of the amino acid sidechains, determines the relative stability of the two helical conformations under a given set of experimental conditions.

The primary interest in peptaibol antibiotics is related to their ability to induce voltage-dependent changes in conductance across artificial bilayer membranes (13, 14). Transitions between the two helical forms may be responsible, in part, for the sensitivity to voltage seen with pores formed by peptaibol antibiotics such as alamethicin (13, 14). One motivation to consider the 3_{10} -helix is the activity of emerimicin on bilayers. Its length of 15 residues is insufficient to span the dielectric thickness of the membrane in the α -helical conformation ($15 \times 1.5 \text{ \AA}$ per residue = 22.5 \AA), while the 3_{10} -helical conformation ($15 \times 2.0 \text{ \AA}$ per residue = 30 \AA) is sufficient. The work of Tosteson *et al.* (15) showing that a 22-residue segment of the S4 repeat of the sodium channel is capable of forming voltage-gated channels stimulates consideration of the 3_{10} -helix as the active form, as positively charged Lys or Arg occurs at every third residue. A bundle of four amphipathic α -helices has been suggested as a plausible structure for the pore-forming element of voltage-gated channels (16). Transition between the α - and 3_{10} -helix could change the relative orientation of sidechains at the channel interface, increasing the apposition of Lys and Arg sidechains, which could increase the diameter of the pore by electrostatic repulsion. A helical transition with the associated increase in length per residue of 0.5 \AA should also be considered as a possible transduction mechanism for transmembrane signaling in receptors which have a single transmembrane segment, such as the insulin receptor.

Early conformational studies of MeA (4), indicated that the available ϕ and ψ values were severely restricted. Further work by Burgess and Leach (5) suggested that MeA was "an obligatory α -helical amino acid residue." These reports considered only the isolated residue. Paterson *et al.* (8) determined the conformational properties of short MeA oligomers, and the influence of asymmetry at the α -carbon atom. The helix-forming ability of poly-(MeA) was investigated by Prasad and Sasisekharan (7) using both the Buckingham and Kitaigorsky potential functions, electrostatics, and a hydrogen-bonding function. The α - and the 3_{10} -helix (17) were recognized as stable species for poly-(MeA), the former being more stable by between 0.3 and 3.6 kcal/mol/residue. Considering recent advances in computational chemistry and the utility of these unusual amino acids in restricting conformational freedom, we felt a reappraisal of these effects using a more satisfactory force-field was necessary.

Methods

SYBYLTM (Tripos Associates, Inc., St. Louis, MO) and AMBER (18) molecular modelling software were used. The Kollman united atom charges for MeA and other

α , α -dialkylamino acids were assigned as previously described (2). Potentials of mean force were calculated using the method of umbrella sampling (19). The distance between the α -carbons of residues 1 and 10 was used as the reaction coordinate, R , and was restrained by a harmonic potential. Ensemble averages were corrected in order to correspond to thermodynamic variables using the method of Valleau and Torrie (20).

For the correlation analysis, 100 oligomers of MeA–Ala were prepared. The length, percent, and substitution pattern for each helix were randomly determined using a simple program which was given the minimum ($n = 4$) and maximum ($n = 18$) helix length and a seed for a pseudo-random sequence. Once the sequences were determined, the models were built and assigned both an α -helical starting conformation and a 3_{10} -helical conformation. Each model was then minimized using the Kollman united atom force field implementation within SYBYLTM to a convergence of 5×10^{-6} kcal with a distance-dependent dielectric of $4r$.

Each pair of helices was evaluated for the resulting energy and conformation. The energetic difference (ΔE) between the 3_{10} -helix and the α -helix was tabulated, and a simple multiple regression analysis (SASTM, SAS Institute, Inc., Cary, North Carolina) was performed on the data using the model in equation 1:

$$\Delta E = c_0 + c_1 L + c_2 P, \quad \text{equation 1}$$

where L is length of the helix in residues and P is the fractional percent of MeA contained in the oligomer.

Results and Discussion

It was necessary to establish a quantitative theoretical basis for understanding the relative stability of α - and 3_{10} -helical structures. In order to characterize possible helical structures, minimizations of oligomers of (MeA) $_n$ and (Ala) $_n$, where $n = 1$ to 15, were performed (2) in order to separate the intrinsic stability per residue from the end effects. The evaluation of MeA oligomers as a function of length showed an increased stability of approximately 1.94 kcal/mol/residue for the isolated α -helix as compared with the 3_{10} -helix for oligomers of MeA. This compares favorably and refines the estimate of between 0.3 and 3.6 kcal/mol/residue by Prasad and Sasisekharan (7), depending on the potential function. The enthalpic difference between the 3_{10} -helix and the α -helix was -2.36 kcal/mol/residue for similar oligomers of Ala, where previous workers (7) had estimated between -3.3 and -3.6 kcal/mol/residue. These calculations are consistent with the small number of observations of 3_{10} -helix in proteins (1) as well as the increased probability of 3_{10} -helix with increased content of MeA residues.

The correlation analysis (Clark and Marshall, unpublished) defines a dichotomy between the intrinsic enthalpic difference and the end effects by use of length as a correlation parameter. These are referred to as the length-dependent (G_L) and length-independent (I_L) energy differences. The coefficients of the regressed data with the standard deviations in parenthesis are shown in equation 2, where ΔE is the difference in energy between the α - and 3_{10} -helix, P is the functional percent of MeA, and L is the length of the peptide.

$$\Delta E = -10.44(1.28) - 1.79(1.40)*P + 2.37(0.07)*L \quad \text{equation 2}$$

$$r^2 = 0.9294.$$

The percent term has a negative coefficient, indicating that the ΔE decreases when

the fractional percent of MeA is increased within a given peptide (i.e., stabilizing the 3_{10} -helix relative to the α -helix). In addition, length can be seen to cause an increase in the ΔE , thus causing α -helices to predominate at longer lengths. The intercept value is negative (-10.44) and compares favorably to the estimated end effect (2) derived from a previous analysis where I_L varied from -8.6 to -13.5 kcal/mol depending on the choice of dielectric. The source of this length-independent stabilization of the 3_{10} -helix results primarily from an additional hydrogen bond as well as enhanced van der Waals interactions (2).

A 3_{10} -helix forms an additional hydrogen bond for the same length peptide due to the i to $i-3$ rather than the i to $i-4$ hydrogen bonding of the α -helix. Thus, for short peptides, a 3_{10} -helix conformation is favored in solvents of low dielectric. As the peptide length increases, the inherent energetic stability of the α -helix dominates and the α -helix becomes dominant in solution. Our calculations (2) *in vacuo* estimate the difference in end effects between the α - and 3_{10} -helix at a dielectric of 4ϵ to be 8.6 kcal/mol, which is primarily due to the extra hydrogen bond. Modulation of the 3_{10} - to α -helix transition length by variation of the dielectric constant of the solvent should therefore be experimentally observable. While the exact transition length will depend on the solvent and peptide sequence, we can estimate that most nonpolar solvents would require a length longer than seven residues for the α -helix to be favored for poly-MeA. This is supported by the correlation analysis where, if ΔE is allowed to be zero (the α - and 3_{10} -helices are equally stable), for $P = 1.0$, the associated length is 6 residues (rounding up from 5.2).

Evidence from solution NMR studies support helix preference as a function of solvent polarity. Pentapeptides containing repetitive MeA-Ala or MeA-Val sequences showed (21) three intramolecular hydrogen bonds in both CDCl_3 and d_6 -DMSO consistent with 3_{10} -helical structure. Heptapeptides of similar sequence (21) had five hydrogen bonds in CDCl_3 (implying a 3_{10} -helix), but only four hydrogen bonds in the more polar d_6 -DMSO consistent with α -helical conformation. Balaram *et al.* (22) have reported that two decapeptides, Boc-MeA-Val-MeA-MeA-Val-Val-Val-MeA-Val-MeA-OMe and Boc-MeA-Leu-MeA-MeA-Leu-Leu-Leu-MeA-Leu-MeA-OMe, show the presence of eight intramolecular hydrogen bonds in CDCl_3 , but only seven hydrogen bonds in d_6 -DMSO. This is consistent with a transformation from 3_{10} -helix to α -helix upon increasing the polarity of the solvent and decreasing the intramolecular hydrogen bond strength. Recent NMR studies (23) of emerimicin 1-9 benzyl ester in which all resonances have been assigned indicate predominately 3_{10} -helix in d_6 -DMSO, as contrasted with the α -helical structure found in the crystal (12). It is interesting that the solution data for an emerimicin 2-9 fragment, Z-MeA-MeA-MeA-Val-Gly-Leu-MeA-MeA-OMe, were consistent with the presence of a right-handed α -helix (24) in trifluoroethanol, a hydrogen bonding solvent, whereas the peptide has a 3_{10} -helical conformation in the crystal.

In the crystal, intermolecular interactions become important and may actually dominate. The predicted enhanced stability of the α -helix in solution once a critical length is reached stands in contrast to the observation of 3_{10} -helical structure in crystals for numerous oligomers of MeA (25) as well as poly-MeA (26, 27) itself. One factor governing the crystal structures in poly-MeA is an improved packing for the 3_{10} -helix as compared with the α -helix arising from the interdigitation of the methyl sidechains. Another major aspect is electrostatic, as the 3_{10} -helix has a smaller radius (1.9 versus 2.3 Å) than the α -helix and a similar dipole moment, 35.6 versus 36.8 Debye

for the MeA decamer, which would result in a stronger electrostatic interaction during antiparallel helical stacking. The α -helical peptides containing multiple α,α -dialkyl amino acids seen in crystals, therefore, arise when the sequence of residues leads to unfavorable packing of the 3_{10} -helix.

Since the calculations and experimental data both suggest that peptides containing multiple α,α -disubstituted amino acids are capable of assuming both helical forms, it was important to determine more precisely their relative free energies and the transition barrier between the two helices. Geometrical parameters of MeA residues from crystal structures of helical peptides were analyzed to identify a pathway for a transition between the two helical forms. A plot of helical pitch (axial rise per residue) against period (residues per turn) indicated a well-defined reaction coordinate for the transition (figure 1). This transition coordinate was used with umbrella sampling molecular dynamics simulations of $\text{CH}_3\text{CO-MeA}_{10}\text{-NMe}$. Potentials of mean force *in vacuo* describing the internal energy, the Helmholtz free energy and the entropy were generated as a function of helix length, R_0 (Clark, Hodgkin and Marshall, unpublished). The difference between the internal energy, U , of the α -helix minimum (at $R_0 = 14 \text{ \AA}$) and the 3_{10} -helix minimum (when $R_0 = 18 \text{ \AA}$) is 10.4 kcal/mol with a barrier (at $R_0 = 17.5 \text{ \AA}$) 1.5 kcal/mol above the 3_{10} -helix, while the equivalent difference in the Helmholtz free energy, A , is 8.3 kcal/mol , with the barrier 1.3 kcal/mol above the 3_{10} -helix (figure 2). Thus, the 3_{10} -helix is entropically stabilized by 2.1 kcal/mol . These simulations include the intrinsic differences (length-dependent) based on the amino acid residue as well as the length-independent end effects (i.e., the extra hydrogen bond present in the 3_{10} -helix). Further simulations in solvent and with varying peptide

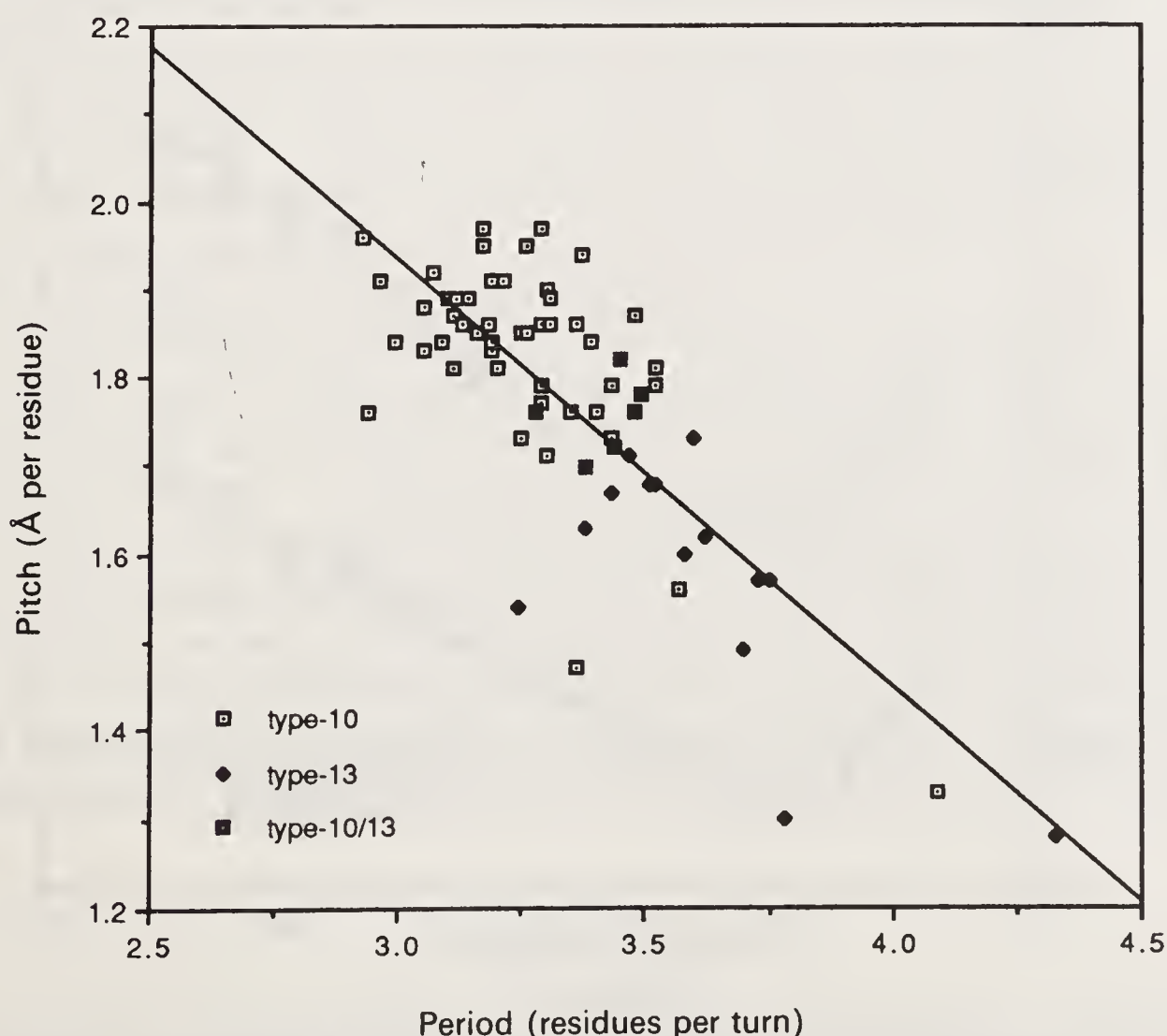


Figure 1. Plot of period vs. pitch for MeA residues in peptide helix crystal structures, coded by the hydrogen-bonding environment.

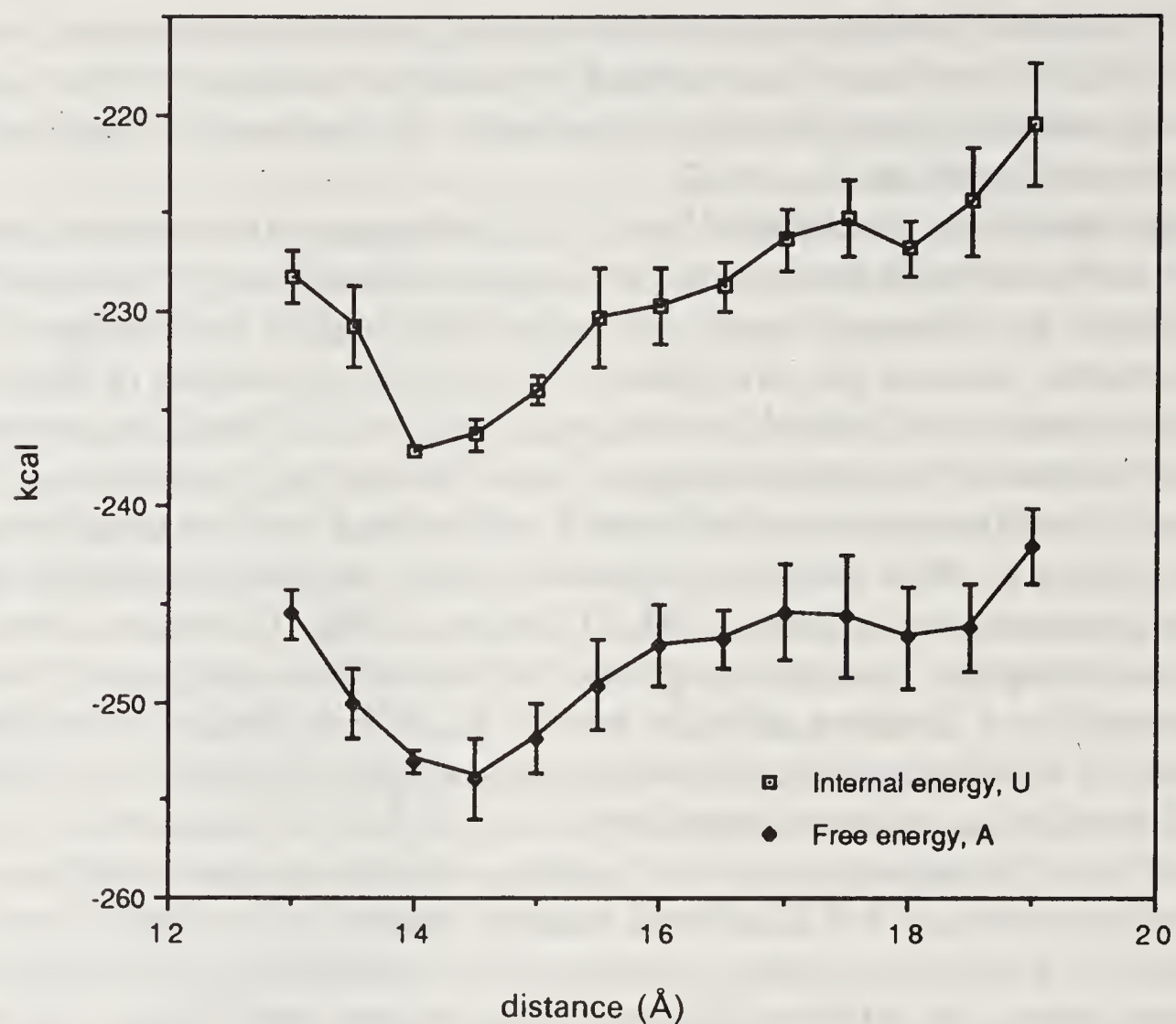


Figure 2. Internal energy, U , and Helmholtz free energy, A , of Ac-MeA₁₀-NMe as a function of helix length. Error bars indicate standard deviations by the method of batch means.

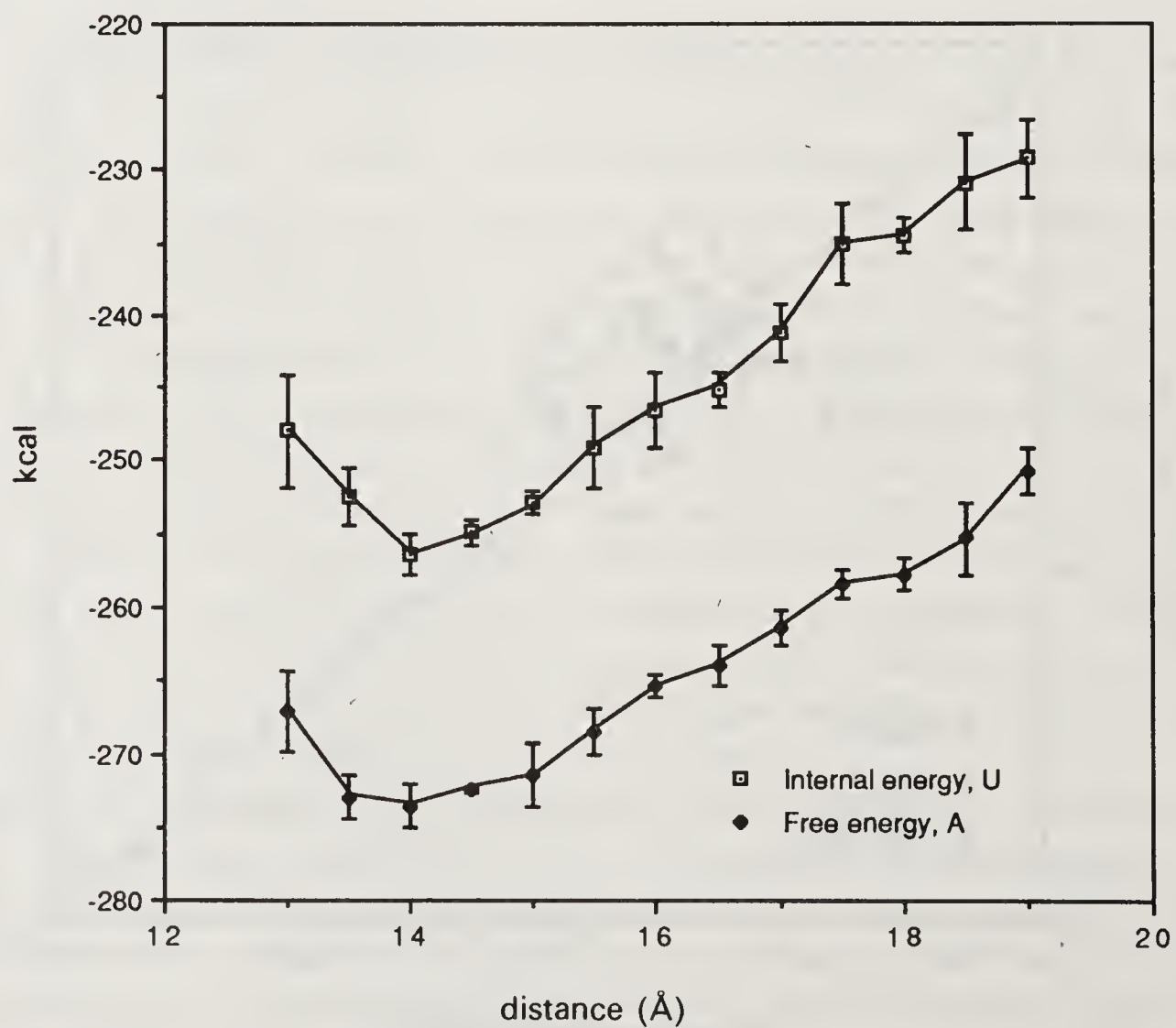


Figure 3. Internal energy, U , and Helmholtz free energy, A , of Ac-Ala₁₀-NMe as a function of helix length. Error bars indicate standard deviations by the method of batch means.

length are necessary to determine the transition length between α - and 3_{10} -helices with a solvent of a given dielectric.

A similar transition coordinate was used with umbrella sampling molecular dynamics simulations of $\text{CH}_3\text{CO-Ala}_{10}\text{-NMe}$. The striking difference between the Ala decamer and the MeA decamer is that the Ala potential has no minima at the 3_{10} -helix (figure 3). This behavior can also be observed by minimizing an Ala-oligomer with a 3_{10} -helix starting conformation; invariably an α -helical final geometry is obtained. The difference between the internal energy, U , of the α -helix minimum (at $R_0 = 14 \text{ \AA}$) and the 3_{10} -helix shoulder ($R_0 = 18 \text{ \AA}$) is 22.0 kcal/mol, while the equivalent difference in the Helmholtz free energy, A , is 15.8 kcal/mol. Thus, the 3_{10} -helix is entropically stabilized by approximately 6.2 kcal/mol; however, the absence of a minima at the 3_{10} -helix as well as a steeper gradient on the potential surface make the 3_{10} -helix much less likely as a congeneric species for poly-Ala. These numbers are somewhat less reliable than those for MeA because of the enhanced flexibility of oligo-Ala and less restricted ϕ, ψ -space, which may offer alternative reaction pathways as well as increase the simulation time necessary to adequately sample configurational space. This may explain the difference between the calculated values where poly-Ala is more entropically stabilized than poly-MeA and the estimates based on potential surfaces (2).

Conclusions

The equilibrium between α - and 3_{10} -helical conformation of peptides containing multiple α, α -dialkyl amino acids depends on length, environment, size and distribution of amino acid sidechains. Once a critical length of seven to eight residues is reached for poly-MeA in solution, the α -helix is favored, especially in more polar solvents. In the crystal, the electrostatic interactions between the dipoles associated with the aligned amide bonds dominate, leading to antiparallel 3_{10} -helices which are associated head-to-tail. As sidechain bulk is increased, the increased effective radius of the 3_{10} -helix causes these electrostatic interactions to become proportionately less of a factor and the inherent stability of the α -helix predominates.

The energy differences between the α -helix and the 3_{10} -helix for decamers of MeA are the same order of magnitude as an additional hydrogen bond. The energetic transition barrier between the two helices is quite low, compatible with the experimental observation of intermediate forms (1, 2) and suggesting that environmental effects, such as solvation, the external electric field and ligand binding, could trigger a conformational transition. While the presence of multiple MeA's decreases the energy difference between the α -helix and the 3_{10} -helix in peptaibol antibiotics and restricts the overall conformations to helical, similar transitions in membrane proteins could be responsible for transduction, as the energy differences between helical types with normal amino acids are well within the range of environmental modulation.

Acknowledgements

This research was supported in part by National Institutes of Health grants GM24483 (GRM) and GM33918 (GRM).

This paper is dedicated to Prof. G. N. Ramachandran.

References

1. Barlow, D. J. & Thornton, J. M. (1988) *J. Mol. Biol.* **201**, 601–619.
2. Hodgkin, E. E., Clark, J. D., Miller, K. R. & Marshall, G. R. *Biopolymers* (in press).
3. Rydel, T. J., Ravichandran, K. G., Tulinsky, A., Bode, W., Huber, R., Roitsch, C. & Fenton, J. W., III (1990) *Science* **249**, 277–280.
4. Marshall, G. R. & Bosshard, H. E. (1972) *Circulation Res.* **30/31** (Suppl. II), 143–150. An earlier preliminary report was published [Marshall, G. R. (1971) *Intra-Sci. Chem. Rep.* **5**, 305–316].
5. Burgess, A. W. & Leach, S. J. (1973) *Biopolymers* **12**, 2599–2605.
6. Pletnev, V. Z., Gromov, E. P. & Popov, E. M. (1973) *Khim. Prir. Soedin.* 224–229.
7. Prasad, B. V. V. & Sasisekharan, V. (1979) *Macromolecules* **12**, 1107–1110.
8. Paterson, Y., Rumsey, S. M., Benedetti, E., Nemethy, G. & Scheraga, H. A. (1981) *J. Am. Chem. Soc.* **103**, 2947–2955.
9. Smith, G. D., Pletnev, V. Z., Duax, W. L., Balasubramanian, T. M., Bosshard, H. E., Czerwinski, E. W., Kendrick, N. E., Mathews, F. S. & Marshall, G. R. (1981) *J. Am. Chem. Soc.* **103**, 1493–1501.
10. Fox, R. O., Jr. & Richards, F. M. (1982) *Nature (London)* **300**, 325–330.
11. Esposito, G., Carver, J. A., Boyd, J. & Campbell, I. D. (1987) *Biochemistry* **26**, 1043–1050.
12. Marshall, G. R., Hodgkin, E. E., Langs, D. A., Smith, G. D., Zabrocki, J. & Leplawy, M. T. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 487–491.
13. Hall, J. E., Vodyanoy, I., Balasubramanian, T. M. & Marshall, G. R. (1984) *Biophys. J.* **45**, 233–247.
14. Menestrina, G., Voges, K. P., Jung, G. & Boheim, G. (1986) *J. Membr. Biol.* **93**, 111–132.
15. Tosteson, M. T., Auld, D. S. & Tosteson, D. C. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 707–710.
16. Montal, M. (1990) *FASEB J.* **4**, 2623–2635.
17. Donohue, J. (1953) *Proc. Natl. Acad. Sci. U.S.A.* **39**, 470–478.
18. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984) *J. Am. Chem. Soc.* **106**, 765–784.
19. Beveridge, D. L. & DiCapua, F. M. (1989) in *Computer Simulation of Biomolecular Systems*, eds. van Gunsteren, W. F. & Weiner, P. K. (ESCOM, Leiden), pp. 1–26.
20. Valleau, J. P. & Torrie, G. M. (1977) in *Statistical Mechanics, Part A: Equilibrium Techniques, Modern Theoretical Chemistry*, ed. Berne, B. J. (Plenum Press, New York), Vol 5, Chapters 4 and 5.
21. Vijayakumar, E. K. S. & Balaram, P. (1983) *Tetrahedron* **39**, 2725–2731.
22. Balaram, H., Sukumar, M. & Balaram, P. (1986) *Biopolymers* **25**, 2209–2223.
23. Beusen, D. D., Hutton, W. C., Kotyk, J. J., Zabrocki, J., Leplawy, M. T. & Marshall, G. R. (in press) in *Peptides 1990*, eds. Giralt, E. & Andreu, D. (ESCOM, Leiden).
24. Toniolo, C., Bonora, G. M., Benedetti, E., Bavoso, A., Di Blasio, B., Pavone, V. & Pedone, C. (1983) *Biopolymers* **22**, 1335–1356.
25. Bavoso, A., Benedetti, E., Di Blasio, B., Pavone, V., Pedone, C., Toniolo, C. & Bonora, G. M. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1988–1992.
26. Malcolm, B. R. (1977) *Biopolymers* **16**, 2591–2592.
27. Malcolm, B. R. (1983) *Bio-polymers* **22**, 319–322.

The Fully Extended Polypeptide Conformation

Claudio Toniolo* and Ettore Benedetti**

* Biopolymer Research Centre, C.N.R., Department of Organic Chemistry, University of Padova, 35131 Padova, Italy;

** Department of Chemistry, University of Napoli, 80134 Napoli, Italy

Introduction

The *fully* extended polypeptide conformation was proposed at an early stage in structural studies of fibrous proteins. In this form H-bonding takes place between the N–H groups of one chain and the C=O groups of the chains on either side, thus making a *planar* sheet held together by *intermolecular* H-bonds directed approximately perpendicular to the chain axis. Neighbouring sheets are then held together by van der Waals forces. In 1951 Pauling and Corey (1) investigated the possibility of small contractions of the polypeptide chains and proposed precise conformations for parallel and antiparallel *pleated-sheet* β -forms which better satisfy stereochemical and H-bonding requirements and have chain repeat lengths nearer those found experimentally. These authors were also able to show that steric hindrance *between adjacent chains* prevents the onset of the planar sheet in case the side chain is anything but a hydrogen, that is, it could be formed only by poly(Gly)_n.

The repeating motif of the fully extended polypeptide conformation is the 2→2 intramolecularly H-bonded form (figure 1). The relative disposition of the two dipoles, N(2)–H(2) and C'(2) = O(2), is such that there is obviously some interaction between them. These four atoms together with the C α (2) atom are involved in a pentagonal cyclic structure, and it is for this reason that this conformation is also called the C₅ structure. The C₅ form has been considered in conformational energy calculations and its occurrence in apolar non-interacting solvents has been proposed using mainly IR absorption and ¹H NMR measurements of model peptides (for review articles see refs. 2 and 3). Gly derivatives have the highest population of C₅ structure if compared to the derivatives of residues carrying a side chain. The influence of the bulkiness of the lateral substituent can easily be explained by considering the *intramolecular* nonbonded

Abbreviations used: Aib, α -aminoisobutyric acid (C α,α -dimethylglycine); Iva, isovaline (C α -methyl, C α -ethylglycine); Deg (C α,α -diethylglycine); Dpg (C α,α -di-*n*-propylglycine); D ϕ g (C α,α -diphenylglycine); Db₂g (C α,α -dibenzylglycine); Ac_nc, 1-aminocycloalkane-1-carboxylic acids; Ac, acetyl; mClAc, monochloroacetyl; Tfa, trifluoroacetyl; Z, benzyloxycarbonyl; *p*BrBz, *para*-bromobenzoyl; OMe, methoxy; OtBu, *tert*-butoxy; NHMe, methylamino; DBH, N', N'-dibenzylacylhydrazido; DMSO, dimethylsulphoxide; TEMPO, 2,2,6,6-tetramethylpiperidinyl-1-oxy.

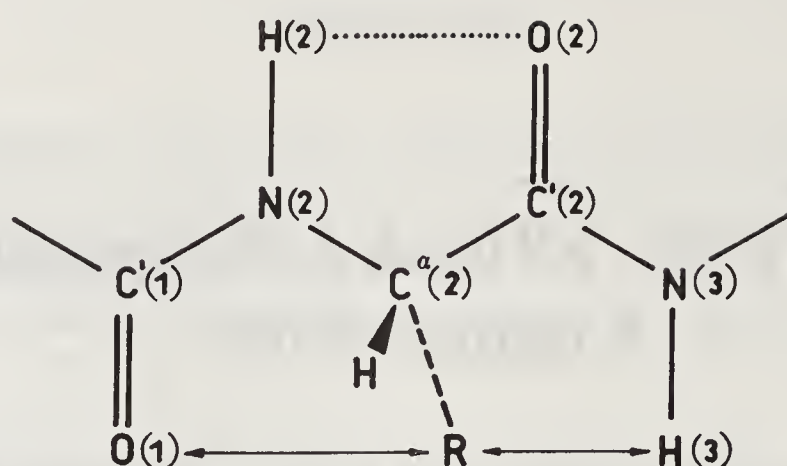


Figure 1. Intramolecular steric repulsions that induce the warping of the $2 \rightarrow 2$ intramolecularly H-bonded (C_5) peptide conformation.

interactions between the group R and the atoms H(3) and O(1), that induce a warping of these asymmetric molecules (figure 1). Unequivocal verification of the occurrence of the C_5 form has been obtained in the crystal-state by the X-ray diffraction analysis of favourable compounds, *i.e.* three Gly- and one L-Ala-containing peptides (2, 4). In a typical example, $^+H_2(Gly)_3O^-$, each NH group shows *bifurcated* interactions, taking part to an intramolecular contact with the carbonyl oxygen following in the chain (C_5 form) and an intermolecular contact with the carbonyl oxygen of an adjacent molecule (4).

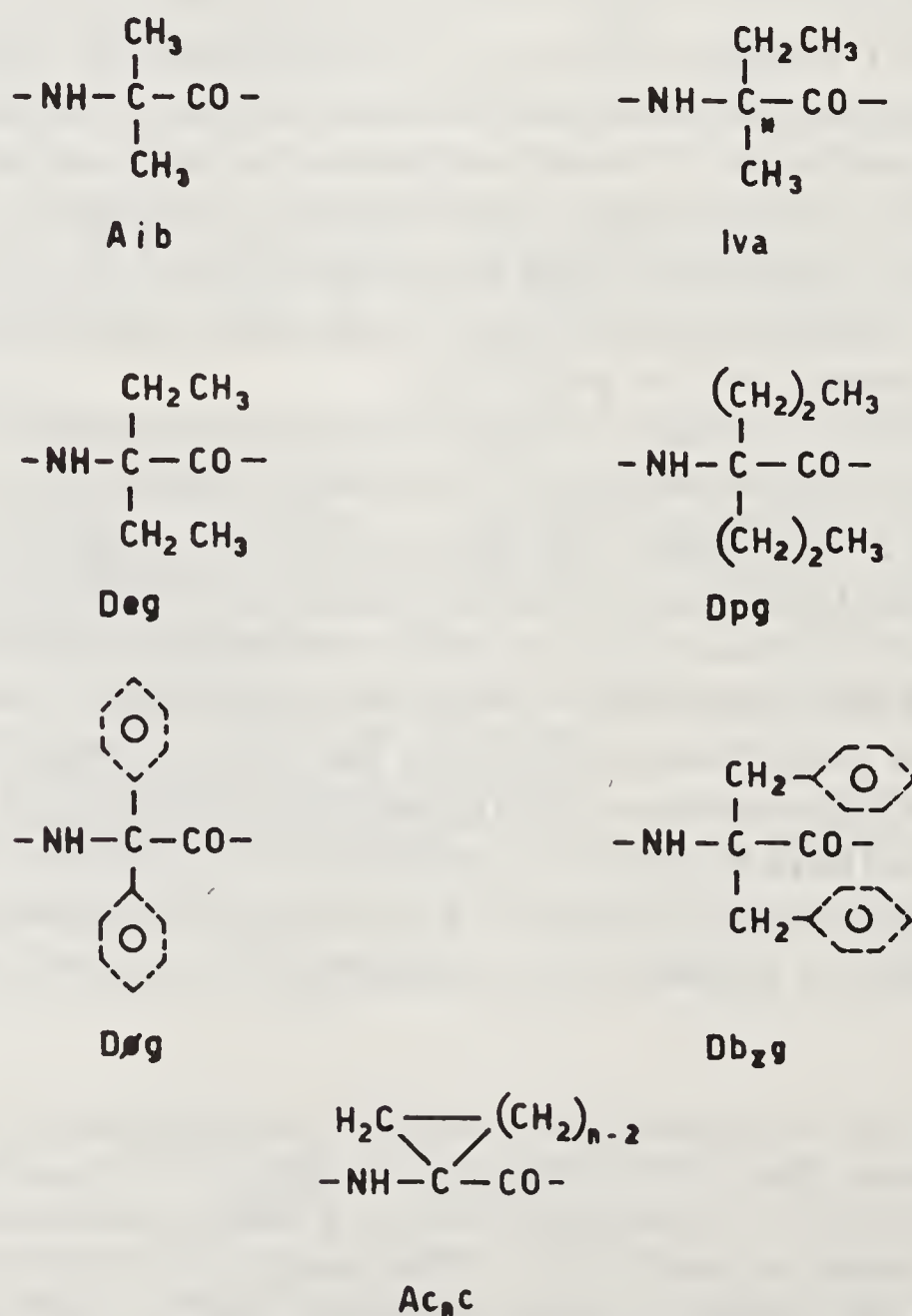


Figure 2. The $C^{\alpha, \alpha}$ -disubstituted α -amino acids discussed in this work.

In this article we wish to summarize the results of our recent conformational energy computations and crystal-state (X-ray diffraction) and solution (IR absorption and ^1H NMR) structural analyses of a variety of derivatives and short peptides from the $\text{C}^{\alpha,\alpha}$ -disubstituted α -aminoacids Aib, Iva, Deb, Dpg, $\text{D}\phi\text{g}$, Db_2g , and Ac_nc (figure 2), that allowed us to characterize the fully extended (C_5) conformation in great detail (5–14).

Results and Discussion

Conformational Energy Computations

The preferred conformations of Ac-X-NHMe [$\text{X} = \text{Aib}$, $\text{D}(=\text{R})\text{-Iva}$, Deg , Dpg , $\text{D}\phi\text{g}$, Db_2g , and Ac_nc] were determined by energy calculations (5, 9, 12–14). The conformational space of these mono-peptides is sterically restricted and low-energy conformations are found only in the regions of the $\alpha/3_{10}$ -helical and fully extended (C_5) structures. As an example, in figure 3 the energy map of Ac-Deg-NHMe is illustrated.

The conformational energies were computed for the helical and the C_5 structures of Aib, Deg, and Dpg mono-peptides as a function of the $\text{N-C}^{\alpha}\text{-C}'$ bond angle (τ). For the Aib mono-peptide, in the wide range of τ values considered, the helical structures remain the most stable, although the energy difference decreases on decreasing the τ angle. By contrast, for the Deg and Dpg mono-peptides the stability order of the helical and C_5 conformations depends markedly on the τ angle, the C_5 conformation being favoured for $\tau \leq 107^\circ$. From our analysis it was also possible to establish that bulky substituents at the C^{α} atom induce a small τ angle, which, in turn, favours the C_5 conformation.

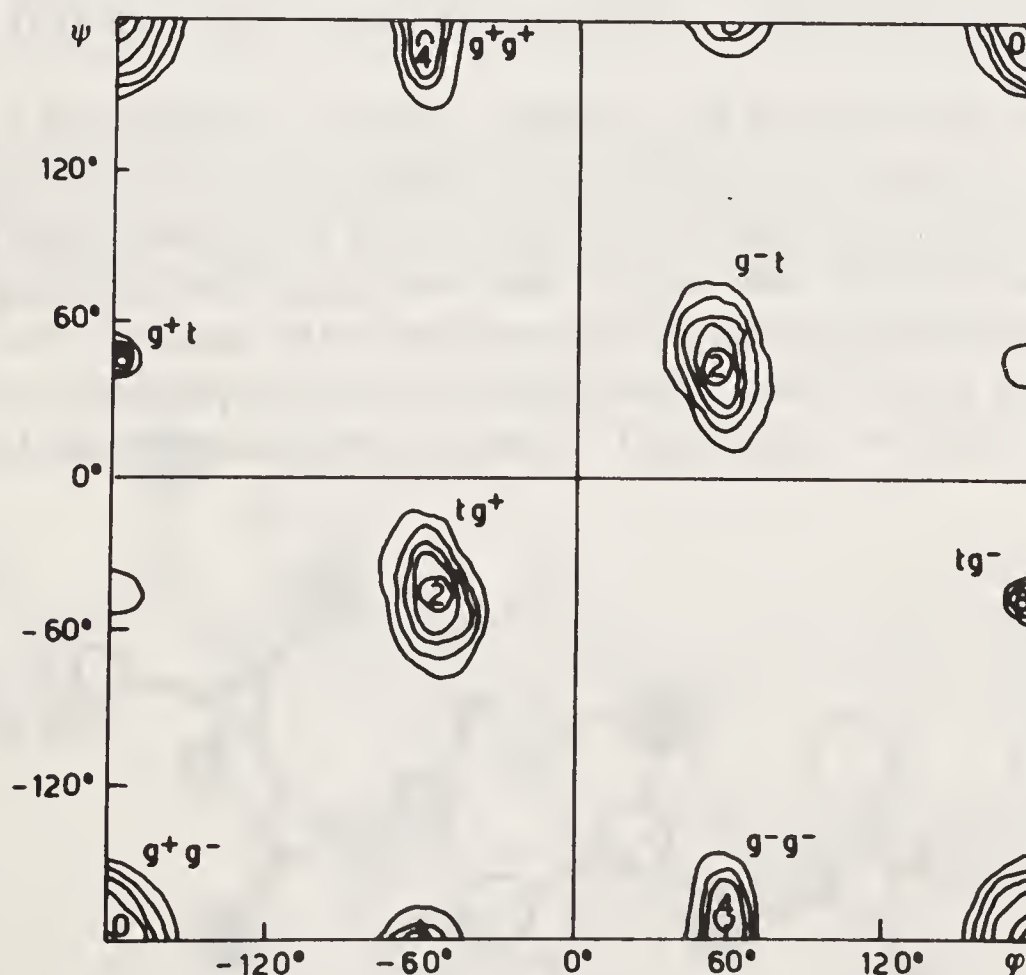


Figure 3. Conformational energy contour map for Ac-Deg-NHMe . The contour lines are spaced 1 kcal/mol over the 10-grid of lowest energy, located at $(\phi, \psi) = (180^\circ, 180^\circ)$. The preferred side-chain conformation in each region is specified. The value used for the τ bond angle is 103° .

Not surprisingly, the behaviour of the chiral Iva mono-peptide is intermediate between those of Aib and Deg, in the sense that the energies of the helical and C_5 conformations are comparable. In addition, the energy minimum for $D\phi g$ and Db_zg mono-peptides is found in the C_5 region, while that for the $Ac_n c$ mono-peptides is in the helical region.

This study has demonstrated *inter alia* the sensitivity of conformational preferences to the geometry and has determined a connection between the narrowing of the τ angle (induced by bulky substituents at the C^α atom) and the occurrence of the C_5 conformation. The results of our systematic and detailed theoretical analyses agree well with those published by other authors, who focused their investigations mainly on Aib peptides (15–20).

Crystal-State Studies

In an effort to ascertain unambiguously the molecular structure of the derivatives and peptides of Aib, Iva, Deg, Dpg, $D\phi g$, Db_zg , and $Ac_n c$ residues in the solid state, the extremely high propensity of these compounds to crystallize was exploited for extensive X-ray diffraction analyses.

Among the vast body of Aib, Iva, $Ac_n c$ compounds examined to date, only two examples have been found to adopt the C_5 conformation. These compounds are $H-(Aib)_2-OtBu$ (C-terminal residue) (8) (figure 4) and $mClAc-D-Iva-OH$ (21) (the structure of the latter was solved by Jung and coworkers). In both cases the C_5 form is additionally stabilized by an intramolecular H-bond, between the peptide NH and the amino nitrogen in $H-(Aib)_2-OtBu$ and between the amide NH and the chlorine atom in $mClAc-D-Iva-OH$. Globally, the two molecules exhibit a flat backbone structure, characterized by a three-centre, doubly intramolecular (bifurcated) H-bonded, fused “ C_5 ”, C_5 form (2).

In contrast to Aib, Iva and $Ac_n c$ residues, the C_5 conformation is the common observation for the Deg, (7, 9, 22) and Dpg (5, 7) residues in the crystal state. From these studies it appears that in 19 (out of 21) residues of the $C^{\alpha,\alpha}$ -symmetrically disubstituted α -aminoacids with linear side chains Deg and Dpg characterizing ten different derivatives and terminally-blocked homopeptides (to the pentamer level) the sets of ϕ , ψ angles are close to 180° , 180° . Interestingly, these homopeptides represent the first example, apart from the $^+H_2(Gly)_3O^-$ tripeptide mentioned in the Introduction, in

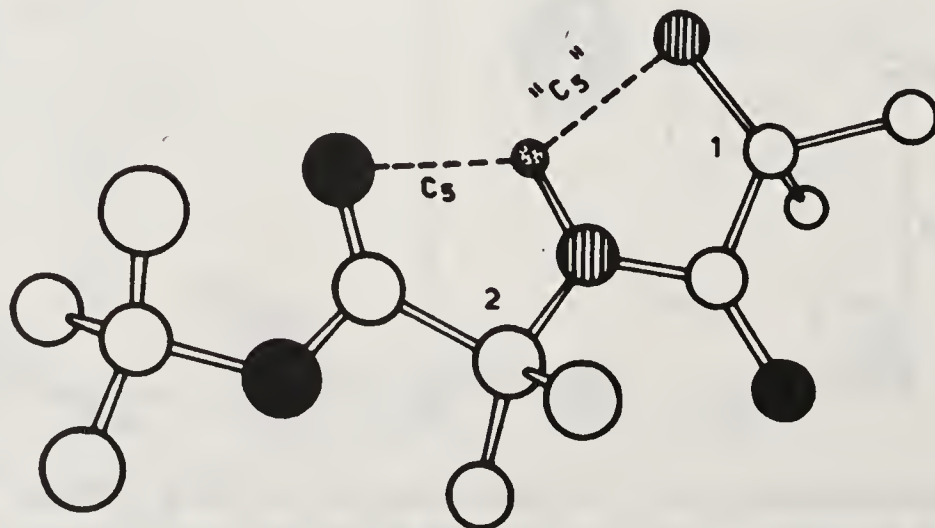


Figure 4. Molecular structure of $H-(Aib)_2-OtBu$. The intramolecular H-bonds are represented as dashed lines.

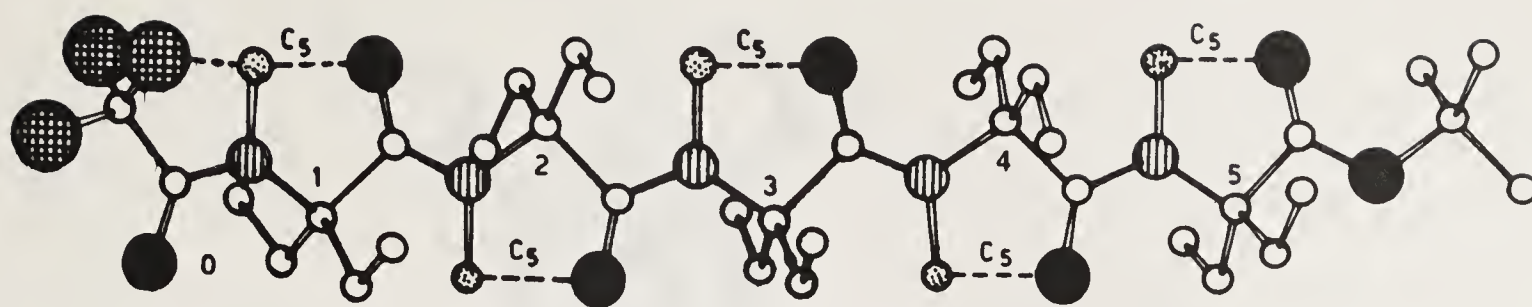


Figure 5. Molecular structure of Tfa-(Deg)₅-OtBu. The intramolecular H-bonds are represented as dashed lines.

which *consecutive* C₅ forms (the fully extended polypeptide conformation) have been experimentally observed. Again, in the α -haloacylated derivatives and peptides the N-terminal C₅ form may be stabilized by an additional N-H...X (halogen) intramolecular H-bond. Illustrative examples for the Deg and Dpg series are shown in figures 5 and 6, respectively. For the first time, in Tfa-(Deg)₃-OtBu, a crystallographically planar peptide skeleton for the main chain atoms ($\phi = \psi = \omega = 180^\circ$) has been observed; in fact, the molecules lie on the mirror plane and the two side chains of each residue are each the mirror image of the other. The ethyl and *n*-propyl side chains of Deg and Dpg residues, respectively, are also fully extended, in order to relieve the unfavourable intramolecular side-chain to main-chain and side-chain to side-chain interactions. It is also worth noting that the N-H and C=O groups characterizing the consecutive C₅ forms are *not* involved in the *intermolecular* H-bonding schemes.

This set of compounds allowed us to characterize the C₅ conformation in detail for the first time. The average parameters are presented in figure 7. In particular, the bond angles show a clear-cut trend: the angles internal to the pentagonal ring tend to be smaller, while those involving atoms of the main chain external to the ring system tend to be larger than the corresponding average angles observed in peptides (23). Interestingly, the critical τ bond angle is narrowed to $\approx 103^\circ$.

More recently, we have extended the structural analysis to D ϕ g- and Db₂g-containing compounds. Representative molecular structures are illustrated in figures 8

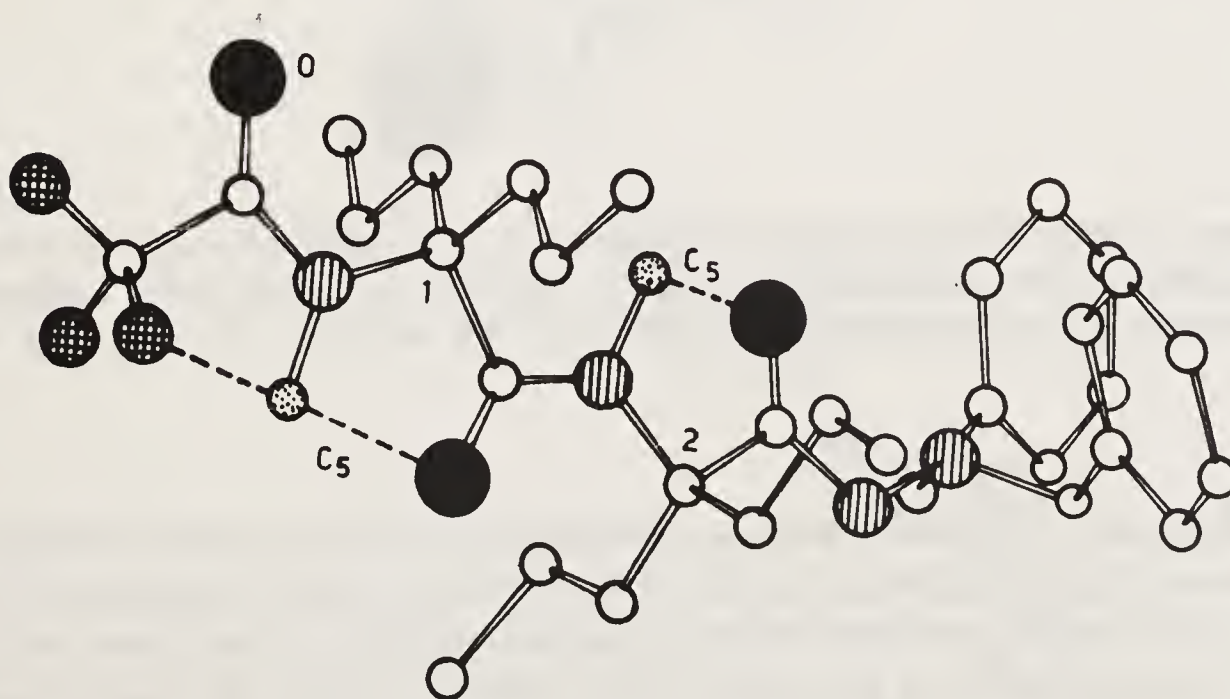


Figure 6. Molecular structure of Tfa-(Dpg)₂-DBH. The intramolecular H-bonds are represented as dashed lines.

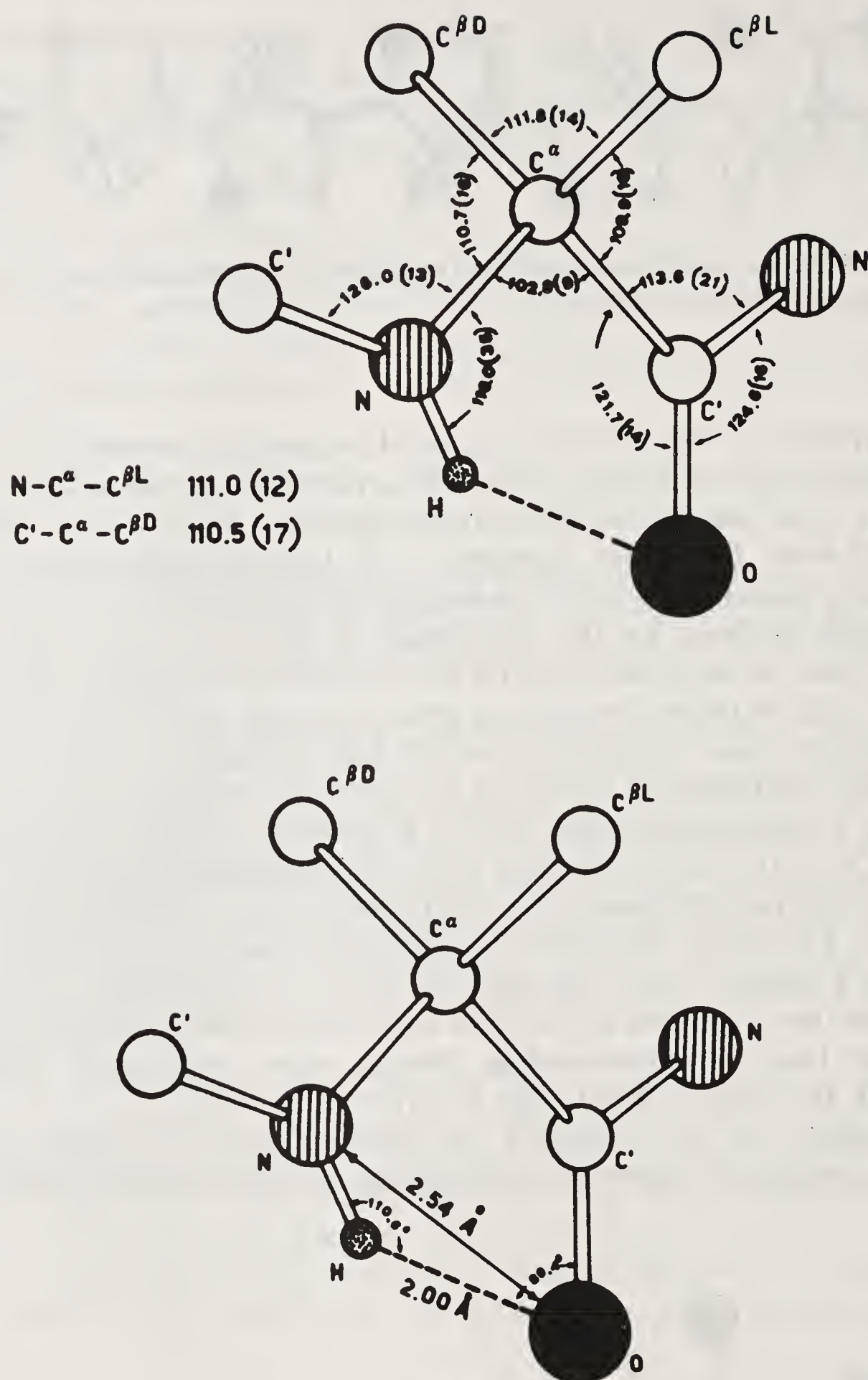


Figure 7. Geometrical parameters characterizing the intramolecularly H-bonded C_5 structure. The values shown are the average from X-ray diffraction analyses of derivatives and homopeptides of the Deg and Dpg residues to the pentamer level.

and 9, respectively. The conformational preferences of these aromatic residues tend to parallel those of their aliphatic analogues Deg and Dpg. In fact, seven (out of nine) $D\phi g$ (12, 24) and all five Db_{zg} (13) derivatives and short peptides investigated so far adopt a typical C_5 conformation in the crystal state. Here too (with the single exception of $mClAc-Db_{zg}-OH$ that shows only the C_5 structure) the fused “ C_5 ”, C_5 form is characteristically seen when the N-terminal $D\phi g$ or Db_{zg} residue is α -haloacylated.

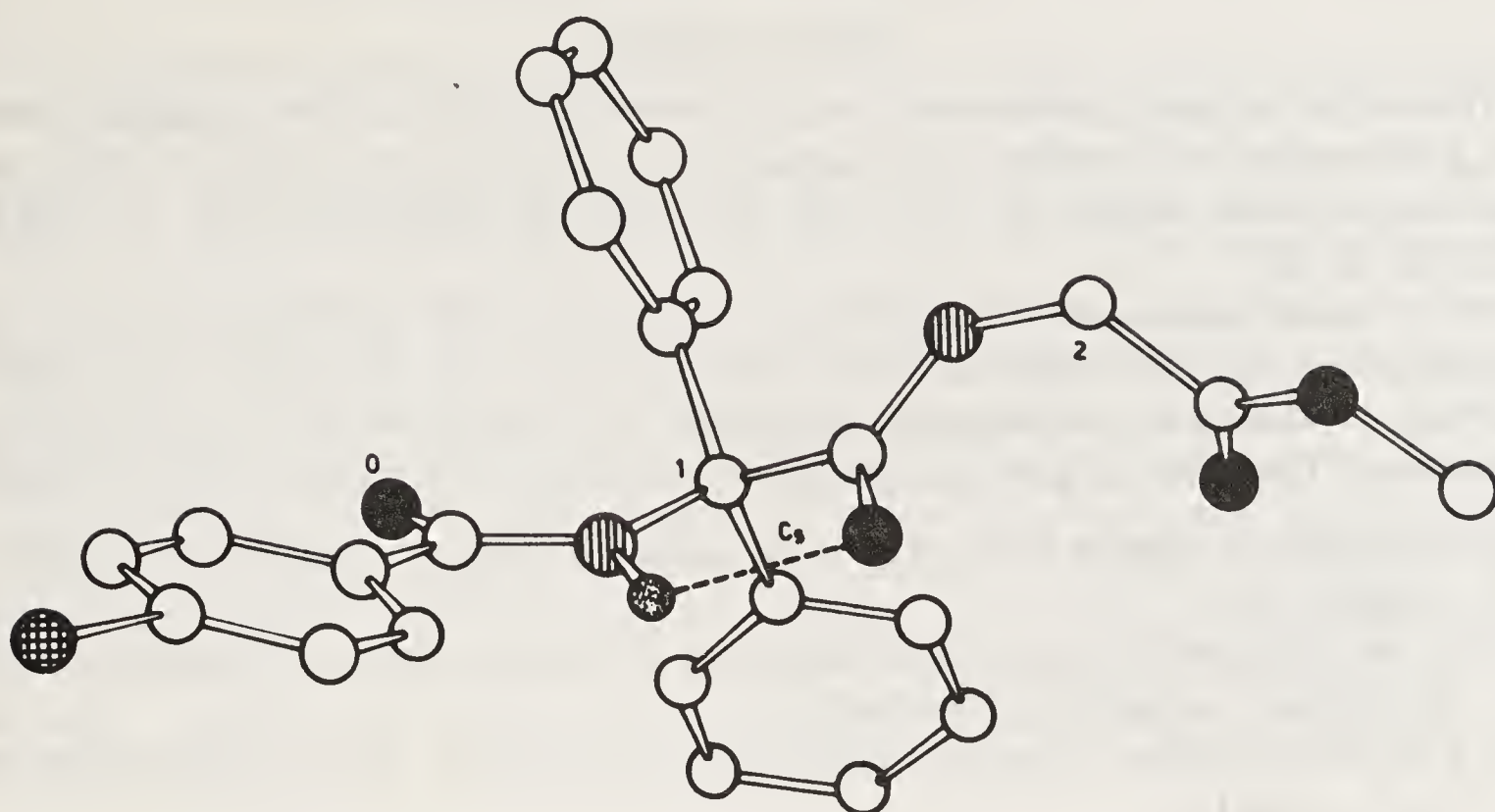


Figure 8. Molecular structure of *p*BrBz-D ϕ g-Gly-OMe. The intramolecular H-bond is represented as a dashed line.

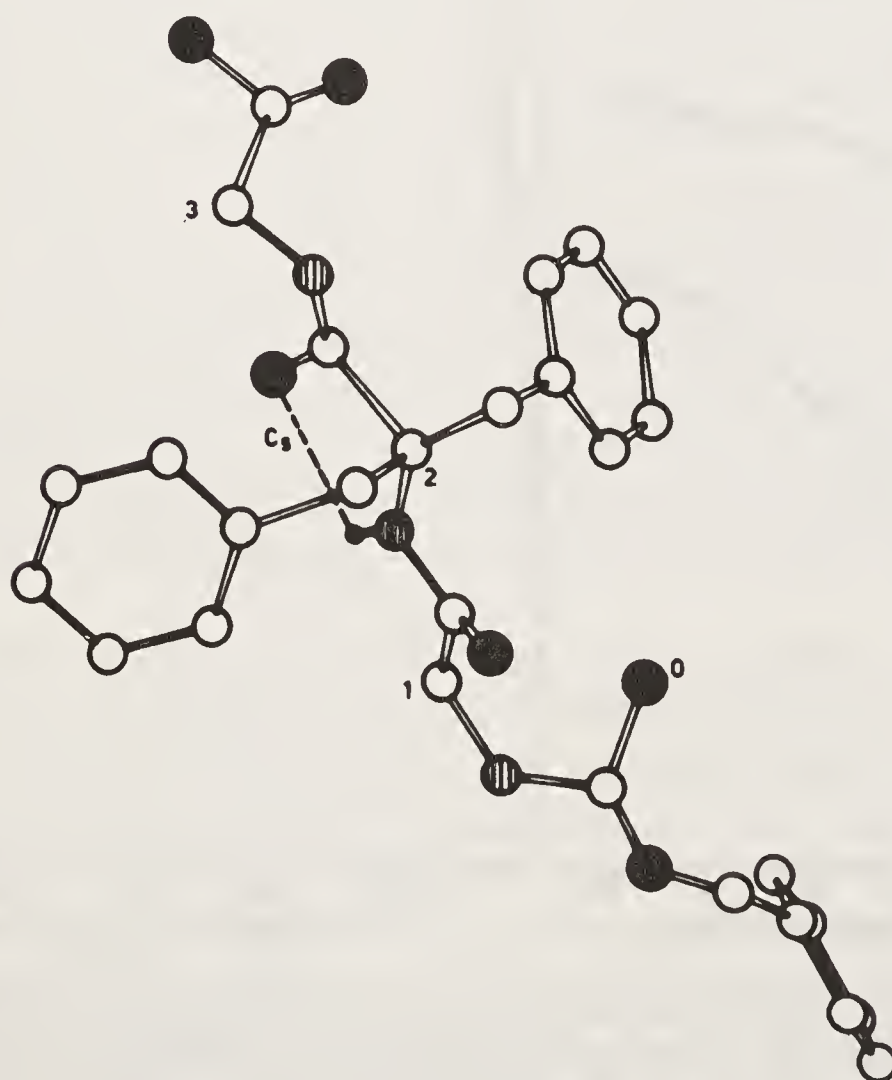


Figure 9. Molecular structure of Z-Gly-Db_{zg}-Gly-OH. The intramolecular H-bond is represented as a dashed line.

Solution Studies

The conformational preferences of selected, terminally-blocked Deg, Dpg, D ϕ g, and Db_zg derivatives and peptides were examined in a solvent of low polarity (CDCl₃) in the concentration range 10^{-2} – 10^{-4} M by using IR absorption and ¹H NMR (6, 7, 10, 12, 13).

In the most informative 3500–3250 cm⁻¹ region the IR absorption spectra are characterized by two bands at 3445–3430 cm⁻¹ (free N–H groups) and 3400–3350 cm⁻¹ (weakly intramolecularly H-bonded N–H groups of the C₅ type) (results not shown). The latter assignment is justified on the basis of the following observations:

- (i) No significant change is observed in the spectra by changing derivative or peptide concentration.
- (ii) In the N^α-blocked amino acid esters the C₅ form is the only possible intramolecularly H-bonded conformation.
- (iii) The 3400–3350 cm⁻¹ spectral region is that expected for the absorption band of the C₅ form (3).

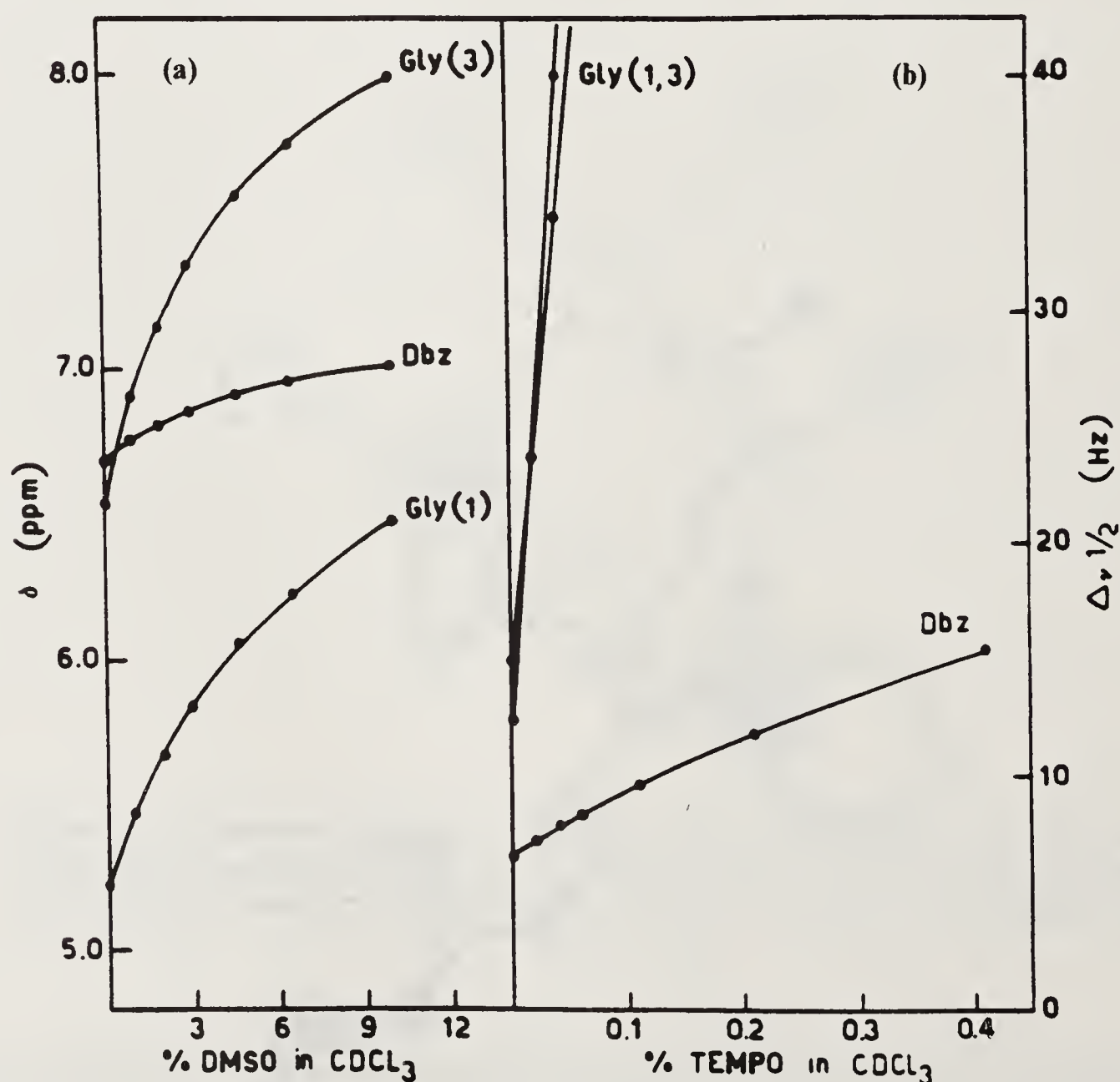


Figure 10. (a) Plot of NH chemical shifts in the ¹H NMR spectra of Z-Gly-Db_zg-Gly-OtBu as a function of increasing percentages of DMSO in CDCl₃ solution (v/v). (b) Plot of the bandwidth of the NH protons of the same peptide as a function of increasing percentages of TEMPO (w/v) in CDCl₃. Conc. 1.1×10^{-3} M.

- (iv) The spectra of the homo-peptides seem to be only slightly sensitive to main-chain elongation.

In order to delineate NH protons solvent accessibilities the effect on NH resonances of heating and addition of DMSO and TEMPO to the CDCl_3 solution was examined in the ^1H NMR spectra. The effect of the two perturbing agents on the NH resonances of a representative peptide, Z-Gly- Db_2g -Gly-O t Bu, is shown in figure 10. The ^1H NMR data confirm the IR absorption results discussed above in the sense that the chemical shifts of the NH protons of all $\text{C}^{\alpha,\alpha}$ -disubstituted residues are concentration independent and exhibit remarkably small variations as the main chain is elongated (in the case of the homo-peptide series), the temperature is raised, and DMSO and TEMPO are added. On the basis of all these findings we are inclined to conclude that in CDCl_3 solution the incorporation of these $\text{C}^{\alpha,\alpha}$ symmetrically disubstituted residues strongly favours the onset of (consecutive) intramolecularly H-bonded C_5 conformations.

Conclusions

The results of the present theoretical and experimental (crystal-state and solution) investigation strongly support the view that Deg, Dpg, $\text{D}\phi\text{g}$, and Db_2g residues, with $\text{C}^{\alpha,\alpha}$ -symmetrically disubstituted side chains, tend to adopt the fully extended, intramolecularly H-bonded C_5 -ring structure, in the $(180^\circ, 180^\circ)$ region of the (ϕ, ψ) conformational map. The agreement between the preferred conformation calculated *in vacuo* and that found in the crystal state and in a solvent of low polarity highlights a remarkable stability for this structure. From our extensive X-ray diffraction analysis we have been able to characterize this polypeptide conformation at atomic resolution. Interestingly, in contrast to peptides rich in Gly, a *symmetrical, unsubstituted* α -aminoacid residue, that favours a fully extended conformation additionally stabilized by *intermolecular* H-bonds between antiparallel chains (4), in peptides rich in the aforementioned *symmetrical, bulky* residues, the N-H and C=O groups involved in the intramolecularly C_5 form are *not* implicated in the intermolecular H-bonding schemes, probably because steric hindrance prevents adjacent chains to approach sufficiently. In conclusion, it seems safe to assume that the incorporation of Deg, Dpg, $\text{D}\phi\text{g}$, and Db_2g residues into bioactive linear peptides would result in the marked stabilization of the C_5 form.

From a comparison of the studies of Aib, Iva, and Ac_nc peptides, where it has been shown that these residues strongly prefer $3_{10}/\alpha$ -helical structures, and with the studies of the Deg, Dpg, $\text{D}\phi\text{g}$, and Db_2g peptides, where it has been shown that these residues markedly favour the fully extended (C_5) conformation, it is evident that the latter structure becomes more stable than the helical structures when the *two* side-chain C^β atoms are *symmetrically* substituted but *not* interconnected in a cyclic system. This conclusion is indirectly supported by our recent observation that the behaviour of C^α -methylated *chiral* α -aminoacids (25) parallels that of their prototype Iva residue. To shed further light on this structural problem we are currently investigating the preferred conformation of chiral residues with the two side-chain C^β atoms unsymmetrically substituted.

Acknowledgements

The authors wish to thank their Italian colleagues at the Biopolymer Research Centre, C.N.R., Department of Organic Chemistry, University of Padova (G. Valle, M. Crisma, G. M. Bonora), at the Department of Chemistry, University of Napoli (C. Pedone, B. Di Blasio, V. Pavone, V. Barone), and at the Institute of Chemistry, University of Basilicata, Potenza (F. Lelj, A. Bavoso) for their active contribution given to this program. Additional stimulating collaborations with the groups headed by P. M. Hardy (Department of Chemistry, University of Exeter, Exeter, U.K.), M. T. Leplawy (Institute of Organic Chemistry, Technical University, Lodz, Poland), G. Jung (Department of Organic Chemistry, University of Tübingen, Tübingen, Germany), P. Balaram (Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India), and H. L. S. Maia (Centre of Pure and Applied Chemistry, University of Minho, Braga, Portugal) are gratefully acknowledged.

This paper is dedicated to Prof. G. N. Ramachandran.

References

1. Pauling, L. & Corey, R. B. (1951) *Proc. Natl. Acad. Sci. U.S.A.* **37**, 729–740.
2. Toniolo, C. (1980) *C.R.C. Crit. Rev. Biochem.* **9**, 1–44.
3. Cung, M. T., Marraud, M. & Néel, J. (1972) *Ann. Chim. (Paris)* **7**, 183–209.
4. Srikrishnan, T., Winiewicz, N. & Parthasarathy, R. (1982) *Int. J. Pept. Protein Res.* **19**, 103–113.
5. Benedetti, E., Toniolo, C., Hardy, P., Barone, V., Bavoso, A., Di Blasio, B., Grimaldi, P., Lelj, F., Pavone, V., Pedone, C., Bonora, G. M. & Lingham, I. (1984) *J. Am. Chem. Soc.* **106**, 8146–8152.
6. Bonora, G. M., Toniolo, C., Di Blasio, B., Pavone, V., Pedone, C., Benedetti, E., Lingham, I. & Hardy, P. (1984) *J. Am. Chem. Soc.* **106**, 8152–8156.
7. Valle, G., Bonora, G. M., Toniolo, C., Hardy, P. M., Leplawy, M. T. & Redlinski, A. (1986) *J. Chem. Soc., Perkin Trans. 2*, 885–889.
8. Valle, G., Toniolo, C. & Jung, G. (1987) *Gazz. Chim. Ital.* **117**, 549–553.
9. Benedetti, E., Barone, V., Bavoso, A., Di Blasio, B., Lelj, F., Pavone, V., Pedone, C., Bonora, G. M., Toniolo, C., Leplawy, M. T., Kaczmarek, K. & Redlinski, A. (1988) *Biopolymers*, **27**, 357–371.
10. Toniolo, C., Bonora, G. M., Bavoso, A., Benedetti, E., Di Blasio, B., Pavone, V., Pedone, C., Barone, V., Lelj, F., Leplawy, M. T., Kaczmarek, K. & Redlinski, A. (1988) *Biopolymers*, **27**, 373–379.
11. Toniolo, C. & Benedetti, E. (1988) *ISI Atlas of Science: Biochemistry*, **1**, 225–230.
12. Crisma, M., Valle, G., Bonora, G. M., De Menego, E., Toniolo, C., Lelj, F., Barone, V. & Fraternali, F. (1990) *Biopolymers*, **30**, 1–11.
13. Valle, G., Crisma, M., Bonora, G. M., Toniolo, C., Lelj, F., Barone, V., Fraternali, F., Hardy, P. M., Langram-Goldsmith, A. & Maia, H. L. S. (1990) *J. Chem. Soc., Perkin Trans. 2*, 1481–1487.
14. Toniolo, C., Bonora, G. M., Barone, V., Bavoso, A., Benedetti, E., Di Blasio, B., Grimaldi, P., Lelj, F., Pavone, V. & Pedone, C. (1985) *Macromolecules*, **18**, 895–902.
15. Marshall, G. R. (1971) in *Intra-Science Chemistry Reports*, ed. Kharasch, N. (Gordon and Breach, New York), pp. 305–316.
16. Burgess, A. W. & Leach, S. J. (1973) *Biopolymers* **12**, 2599–2605.
17. Venkataram Prasad, B. V. & Sasisekharan, V. (1979) *Macromolecules* **12**, 1107–1110.
18. Paterson, Y., Rumsey, S. M., Benedetti, E., Némethy, G. & Scheraga, H. A. (1981) *J. Am. Chem. Soc.* **103**, 2947–2955.
19. Venkataram Prasad, B. V. & Balaram, P. (1984) *C.R.C. Crit. Rev. Biochem.* **16**, 307–348.

20. Barone, V., Lelj, F., Bavoso, A., Di Blasio, B., Grimaldi, P., Pavone, V. and Pedone, C. (1985) *Biopolymers* **24**, 1759–1767.
21. Bosch, R., Brückner, H., Jung, G. & Winter, W. (1982) *Tetrahedron* **38**, 3579–3583.
22. Galdecki, Z. and Luciak, B. (1984) *Acta Crystallogr.*, (Suppl.) **A40**, 03·2–17.
23. Benedetti, E. (1982) in *Chemistry and Biochemistry of Amino Acids, Peptides, and Proteins*, ed. Weinstein, B. (Dekker, New York), Vol. 6, pp. 105–184.
24. Eggleston, D. S., Valle, G., Crisma, M. & Toniolo, C., to be published.
25. Valle, G., Crisma, M., Toniolo, C., Polinelli, S., Boesten, W. H. J., Schoemaker, H. E. & Kamphuis, J. (1991) *Int. J. Pept. Protein Res.*, in press.

A Guest–Host Approach to Oligodepsipeptides: Studies of Oligoglutamates and Depsiglutamates in Trifluoroethanol

Y. V. Venkatachalapathi*, R. Katakai[†] and M. Goodman[‡]

Department of Chemistry, University of California, San Diego, La Jolla, California 92093, USA

[†]Department of Chemistry, College of Technology, Gunma University, Tenjin-cho, Kiryu-shi 376, Japan

Hydrogen bonds are an important stabilizing factor in the three-dimensional structures of proteins and peptides. We were involved in the study of partial replacements of peptide bonds with structurally similar ester bonds as a means of eliminating specific hydrogen bond donor residues. The effects of the presence of S-lactic acid residues in polyalanine (1,2) poly γ -methyl-L-glutamate (3) and poly-L-lysine- γ -methyl-L-glutamate copolymers have been reported earlier (3).

Oligo- γ -methyl glutamates bound to polyoxyethylene were studied in our laboratory by ¹H-NMR methods (4). These investigations were carried out in trifluoroethanol using multi decoupler channels. Spectra of heptamers in this series in appropriate solvents indicated the presence of folded conformations which are predominantly α -helical. In this paper we report about the NMR studies of oligoglutamate analogs and oligodepsipeptides with one peptide bond in each molecule systematically replaced by an ester bond. Studies were carried out in trifluoroethanol without the help of multi channel decouplers. We studied analogs of the general form Boc-[Glu(OMe)]_m-Lac-[Glu(OMe)]_n-OMe as well as related parent oligopeptides. The largest peptide we report about in this paper is an undecamer. Our aim in this research is to describe the conformational studies of these peptides in a single solvent system by two different spectral techniques: NMR and circular dichroism (CD).

Methods

Synthesis of Peptides

The depsipeptides were synthesized stepwise from dimethyl glutamate hydrochloride (HCl-H-Glu(OMe)-OMe) as starting material. Two strategies were used:

*Present Address: Department of Medicine and Atherosclerosis Research Unit, UAB Medical Center, Birmingham, AL 35294.

[‡]For correspondence.

In the first synthesis, Boc- γ -methyl-glutamic acid was coupled to dimethyl glutamate hydrochloride followed by deprotection of the amine terminus and repeated cycles to prepare the Boc-peptide of desired length. Couplings were carried out either with dicyclohexylcarbodiimide (DCC) and hydroxybenzotriazole (HOBt) in dimethylformamide (DMF) at room temperature or by the mixed anhydride method using isobutylchloroformate (IBCF) and N-methylmorpholine (NMM) in dry tetrahydrofuran (THF) at -20°C . The hydroxyl group of lactic acid was protected as a benzyl ether. This acid was introduced into the peptide chain using DCC in dichloromethane as solvent. The ester linkage between Boc- γ -methylglutamate and lactic acid was synthesized using DCC in the presence of dimethylaminopyridine (5) in dichloromethane.

Oligopeptides of up to seven residues were prepared by stepwise addition of Boc- γ -methyl glutamic acid to the free amino end of the chain begun with dimethyl glutamate as mentioned above. Further elongation of the chain (from octamer to undecamer) was carried out by fragment condensation using DCC/HOBt in DMF at room temperature. When the products were insoluble in ethyl acetate or chloroform, it was not possible to use conventional washing procedures. In such cases, they were purified by HPLC using reversed phase C18 columns.

In the second strategy the chain was elongated with Nps- γ -methyl glutamate derivatives (NPS = nitrophenylsulfenyl). The intermediates Nps-[Glu(OMe)]_m-OMe ($m = 2$ to 5) were prepared by the usual procedure for peptide synthesis with Nps-Glu(OMe)-OH-DCHA and DCC in chloroform. The NPS-protection of the amine terminus was selected because of its facile removal by treatment with very mild acid which does not harm the ester linkages of the depsipeptide bond.

The formation of the depsipeptide linkage between γ -methyl glutamate and lactic acid was achieved using Nps- γ -methyl glutamate-N-carboxyanhydride (Nps-Glu(OMe) NCA). The Nps-Glu(OMe) NCA derivative of γ -methyl glutamate NCA was allowed to react with lactic acid in pyridine and THF for 2 hours at room temperature. Nps-Glu(OMe)-Lac-OH was obtained in a crystalline state and purified by recrystallization from ethyl acetate.

Condensation of Nps-Glu(OMe)-Lac-OH with oligo- γ -methyl glutamate was achieved using DCC in the presence of N-hydroxysuccinimide (HONSu). The oligomer, Nps-[Glu(OMe)]_n-OMe, was treated with two equivalents of hydrochloric acid in dioxane to give a hydrochloride HCl-[Glu(OMe)]_nOMe, which was coupled to Nps-Glu(OMe)-Lac-OH with the aid of DCC and HONSu in chloroform. The resulting oligodepsipeptide Nps-Glu(OMe)-Lac-[Glu(OMe)]_n-OMe was purified by silica gel column chromatography. For conformational analysis the Nps-protection of these oligodepsipeptides was replaced with a Boc group. The Nps group was removed by hydrochloric acid and the hydrochloride was treated with di-*t*-butyloxycarbonyl dicarbonate in the presence of N-methylmorpholine in DMF as solvent to obtain the Boc depsipeptides.

The purity of peptides was checked by thin layer chromatography in different solvent systems (methanol, chloroform, 5:95, and methanol, chloroform, acetic acid, 5:90:5) and by 360 MHz NMR spectroscopy. Partial 360 MHz NMR spectra for the undecamer oligopeptide and the analogous oligodepsipeptide are shown in figure 1. When the compounds were impure, they were purified either by silica gel chromatography or by preparative HPLC with reversed phase C18 columns using

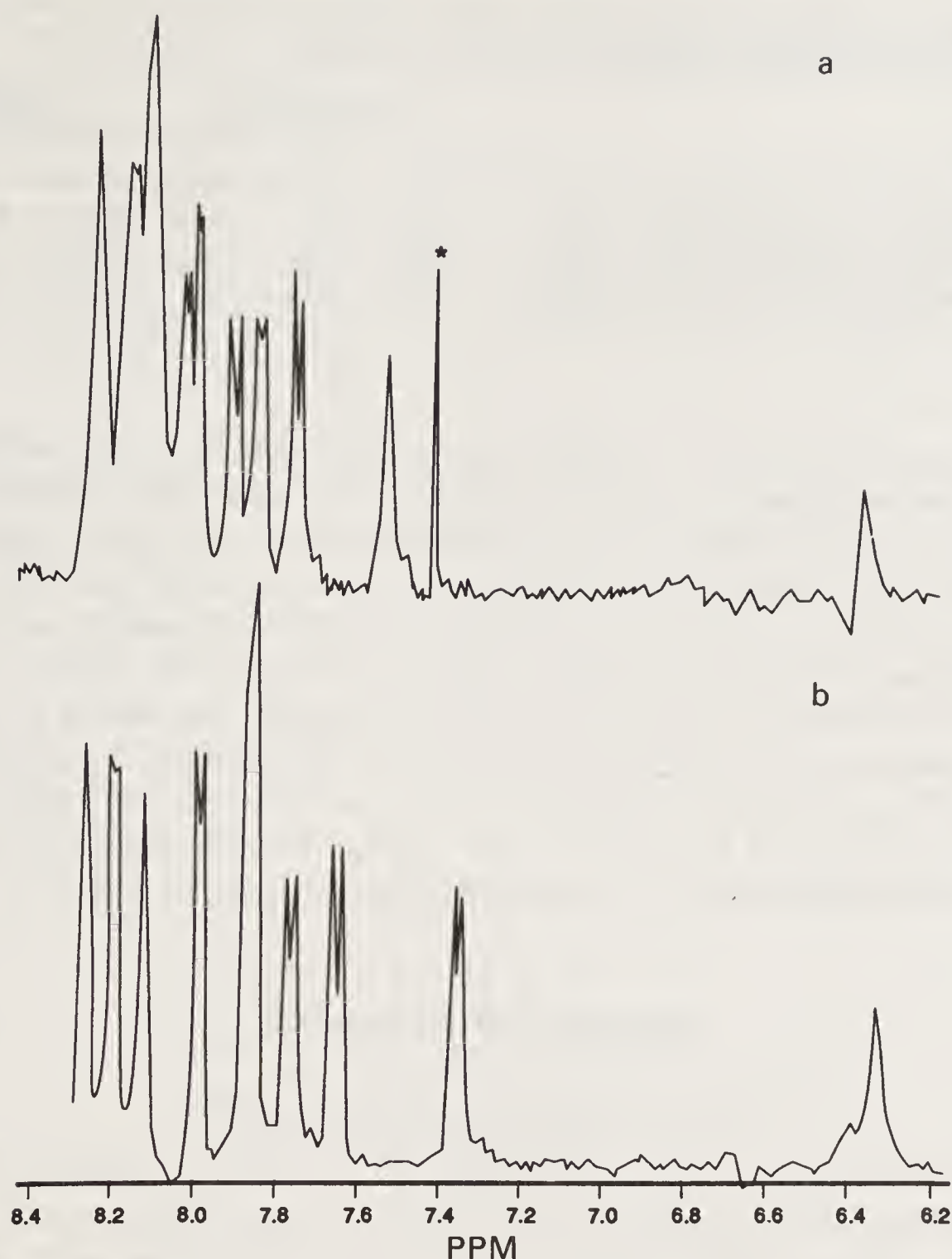


Figure 1. Partial ^1H NMR spectra of (a) Boc-[Glu(OMe)]₁₁-OMe and (b) Boc-[Glu(OMe)]₅-Lac-[Glu(OMe)]₅-OMe in trifluoroethanol. The starred peak is an impurity of CHCl_3 in the solvent.

an isocratic 55% aqueous acetonitrile solvent system. The representative analytical data of some peptides are given in table 1.

Spectral Measurements

Nuclear magnetic resonance (NMR) spectra were obtained on a 360 MHz spectrometer built in-house from a Varian instrument equipped with an Oxford Magnet and Nicolet 1280 Computer. NMR chemical shifts are correlated in ppm from tetramethylsilane (TMS) as the internal standard. The NMR spectra were studied in trifluoroethanol, a solvent known for its helix-promoting properties.

To suppress the hydroxyl signal of trifluoroethanol, the symmetric 1331 pulse sequence was used (6). The methylene quartet was suppressed by irradiation during the delay time between the scans. Since we were mostly interested in the amide proton

Table 1. Elemental analysis of representative depsipeptides.

Compound	Calculated (%)			Found (%)		
	C	H	N	C	H	N
Boc-[Glu(OMe)] ₃ -Lac-[Glu(OMe)]-OMe	51.03	6.75	7.21	51.15	6.71	7.11
Boc-[Glu(OMe)] ₃ -Lac-[Glu(OMe)] ₂ -OMe	50.92	6.68	7.61	51.03	6.78	7.48
Boc-[Glu(OMe)] ₃ -Lac-[Glu(OMe)] ₃ -OMe	50.84	6.74	7.91	50.86	6.74	7.96

resonances, the other option of irradiating the hydroxyl resonance and suppressing the methylene quartet was not chosen. In practice a 50% deuterated solvent was used.

Circular dichroism (CD) measurements were carried out on a Cary 61 spectropolarimeter which was modified by replacing the original Pockel cell with a 50 KHz photoelastic modulator (Hinds International FS-5/PEM-80), used in conjunction with a lock amplifier (EG and G Princeton Applied Research No. 1280) to detect and integrate the modulation. System automation, multiple scan signal averaging and base line subtraction were accomplished by DEC 11/02 computer interfaced directly to both the Cary 61 and the amplifier. The system software and custom hardware interfaces were designed by Allen Microcomputer Services, Inc. CD spectra were obtained with 0.02 cm and 0.1 cm cells, by signal-averaging of 20 scans.

Results and Discussion

NMR Assignments of Oligopeptides

In previous studies (7), the incorporation of α -deuterated methionyl residues at specific sites along the chain (8) was used to assign the amide NH and α -CH resonances to individual amino acids. Similar studies were undertaken with oligo- γ -ethyl glutamates (9). With γ -methyl glutamates bound to polyoxyethylene, the individual amide proton resonances were assigned unambiguously by the sequential replacement of α -deuterated glutamic acid residues at specific sites along the chain (4). Similar patterns of amide resonances were shown to be present in methionyl as well as in γ -methyl glutamyl oligopeptides. The replacements of specific α -amino acid residues with guest α -hydroxy acid residues (lactic acid) allow us to investigate the importance of a specific hydrogen bond in a given oligopeptide conformation (10). NMR studies were carried out in a mixture of 1:1 deuterated and nondeuterated trifluoroethanol and the 1331 pulse sequence was used to suppress the large hydroxyl resonance while the methylene quartet was suppressed by irradiation. It was difficult to observe the α -proton resonances since these were masked under the hydroxyl proton peak generated by the solvent. Consequently, our NMR results are based solely on the resonances of the amide protons.

The assignments of the amide resonances in this study were made using the guest-host approach based on comparison with analogous oligopeptides and the systematic elongation of the peptide chain. In all these compounds the resonances of Glu 1-NH is assigned as that occurring at high field (~ 6 ppm) since this proton belongs to a urethane function. The position of the urethane NH resonance was not studied in detail since it is generally located close to the hydroxyl resonance of

trifluoroethanol and sometimes is also suppressed by the 1331 pulse sequences. Where possible, we included a discussion of the urethane linkage where it was clearly visible in the spectrum.

In Boc-[Glu(OMe)]₇-OPOE and Boc-[Glu(OEt)]₇-OEt the amide resonances were assigned unambiguously by the guest–host approach and by incorporating α -deuterated amino acids at specific sites. A distinct pattern in the positions of the amide NH resonances, observed from low field to high field for residues Glu 2 \rightarrow Glu *n*, respectively, was observed in Boc-[Glu(OMe)]₇-POE peptides (4). These assignments for this oligopeptide conjugate are included in table 2. The amide resonances of the higher members of the series of γ -methyl glutamate oligopeptides are also presented in table 2. It should be noted that the amide proton of the carboxyl terminal residue of the heptamer displays a high field shift (7.55 ppm) since it is covalently linked to the POE chain, whereas those of other residues show a small shift in the opposite direction. The Glu 2-NH resonance in the heptamer exhibits a chemical shift of 8.07 ppm. The octamer shows a chemical shift of 8.13 ppm. This trend is maintained as the chain length of the peptide is increased. This Glu 2-NH resonance appears at 8.17 ppm in the nonamer and moves to 8.23 ppm in the decamer and 8.24 in the undecamer. The chemical shift of the amide proton of the carboxyl terminal residue moves to a higher field with chain elongation. In the octamer it appears at 7.59 ppm, and is found at 7.46 ppm in the undecamer.

The temperature coefficients of the amide resonances of the oligopeptides in TFE are shown in table 2. Based on their temperature coefficients in TFE we can divide the amide protons into two categories: the N-terminal and C-terminal region. In the first group residues 1 and 2 of the heptamer and octamer show temperature coefficients greater than 7 ppb/K. Similarly, in the longer peptides the temperature coefficients of the N-terminal residues are higher than 6 ppb/K. In the nonamer, residues 1, 2 and 3 show values of 7.4, 6.6 and 5.7 ppb/K respectively, and the first four residues

Table 2. Assignment of amide resonances (in ppm) in TFE to individual amino acid residues of oligoglutamates.

Oligopeptide	Chemical shifts and temperature coefficients*										
	Glu 1**	Glu 2	Glu 3	Glu 4	Glu 5	Glu 6	Glu 7	Glu 8	Glu 9	Glu 10	Glu 11
Glu 7***	6.20 (8.3)	8.07 (8.0)	7.93 (4.6)	7.76 (3.8)	7.75 (5.1)	7.61 (3.9)	7.55 (4.6)				
Glu 8	6.24 (8.2)	8.13 (7.5)	7.92 (3.7)	7.86 (4.5)	7.70 (4.2)	7.66 (1.6)	7.66 (4.3)	7.59 (4.5)			
Glu 9	6.26 (7.4)	8.17 (6.6)	8.01 (5.7)	7.86 (4.9)	7.80 (0.1)	7.80 (5.0)	7.74 (4.4)	7.64 (4.6)	7.52 (−0.2)		
Glu 10	6.30 (8.9)	8.23 (6.4)	8.07 (6.8)	8.04 (6.2)	7.94 (4.7)	7.94 (5.9)	7.83 (4.9)	7.79 (−0.4)	7.70 (4.6)	7.47 (−0.2)	
Glu 11	6.30 (7.1)	8.24 (5.9)	8.16 (5.6)	8.16 (5.1)	8.12 (6.2)	8.03 (5.6)	7.97 (4.2)	7.88 (4.5)	7.77 (−0.1)	7.72 (4.1)	7.46 (0.3)

*The values in parentheses are temperature coefficients (ppb/K).

**The resonances in this column arise from the NH of a urethane linkage.

***The heptamer values were obtained from the conjugate with polyoxyethylene, Boc[Glu(OMe)]₇-OPOE.

in the decamer and five residues in the undecamer have similarly higher temperature coefficients (see table 2).

The second group includes the residues in the C-terminal region of the peptides. In the heptamer five residues on the C-terminal end have temperature coefficients between 3.8 and 5.1 ppb/K. Values in this range arise from protons involved in hydrogen bonds at room temperature which become disrupted at higher temperatures (11). Such weak hydrogen bonds can also arise from a C_7 γ -turn.

Temperature coefficients of the C-terminal six residues in the octapeptide were lower than those in the heptapeptide, ranging from 1.6 to 4.5 ppb/K. A temperature coefficient of 1.6 ppb/K is indicative of solvent shielding about a proton most likely involved in an intramolecular hydrogen bond. In the longer peptides the C-terminal residues also show lower temperature coefficients, again indicating the presence of intramolecular hydrogen bonds. The presence of such intramolecular bonds constrains the molecule to prefer ordered structures. In fact, this proposal is supported by circular dichroism data presented later in this paper.

NMR Assignments of Depsipeptides

Table 3 shows the assignments of amide resonances along with the temperature coefficients for the oligodepsipeptides we have studied. In the series Boc-[Glu(OMe)]-Lac-[Glu(OMe)]_n-OMe, when $n = 1$ the assignment of Glu 1 and Glu 3 NH resonances at 5.80 and 7.74 respectively was unambiguous. In the tetrapeptide ($n = 2$) again the resonance at 5.85 ppm is assigned to Glu 1 NH. Among the other two resonances at 7.81 and 7.45 ppm, the former arises from Glu 3 NH whereas the latter belongs to Glu 4 NH. These are in accordance with the oligopeptide studies where the C-terminal amide proton appears at a higher field than the other amide protons in the chain (4). A similar trend of these resonances in chloroform was also noted in our previous studies (10). In the pentapeptide the Glu 1 NH was buried under the hydroxyl resonance of TFE at room temperature. The high field resonance at 7.41 ppm was assigned to Glu 5 while the resonances at 7.93 and 7.53 ppm were due to Glu 3 and Glu 4 NH resonances.

The urethane amide appears at 5.80 ppm whereas the C-terminal amide resonance appears at 7.46 ppm in the hexadepsipeptide ($n = 4$). The remaining three resonances at 8.03, 7.55 and 7.55 ppm were assigned to Glu 3, Glu 4 and Glu 5 amide protons respectively. In the hexa- γ -methyl glutamate bound to POE the same amide protons appear at 7.93, 7.75 and 7.65 ppm respectively. The Glu 3-NH resonance in the depsipeptide is shifted to lower field because of the presence of the adjacent ester bond. The low field shifts of Glu 4 and Glu 5 in the hexadepsipeptides as compared to the same residues in the POE bound hexapeptide are due to their proximity to the carboxyl methyl ester end group. Similar effects were also seen in longer peptides for residues near the free carboxyl group.

In the series Boc-[Glu(OMe)]-Lac-[Glu(OMe)]_n-OMe, the heptadepsipeptide ($n = 5$) urethane amide resonance is buried under the solvent peak at room temperature. A low field shift of the Glu 3 amide proton (8.03 ppm) was seen when compared to the Glu 3 NH in the analogous oligopeptide (7.93 ppm). This shift can be attributed to the inductive effect of the adjacent lactic acid residue. The resonance at 7.51 ppm was assigned to Glu 7 NH which is at higher field than the same proton in the oligopeptide (7.57 ppm). The other three resonances, 7.61, 7.55 and 7.55 ppm,

Table 3. Assignment of amide resonances (in ppm) in TFE to individual amino acid residues of oligodepsiglutamates.

Depsipeptide	Chemical Shifts and Temperature Coefficients*										
	Glu 1**	Glu 2	Glu 3	Glu 4	Glu 5	Glu 6	Glu 7	Glu 8	Glu 9	Glu 10	Glu 11
Boc[Glu(OMe)]Lac[Glu(OMe)]-OMe	5.80		7.74 (8.0)								
Boc[Glu(OMe)]Lac[Glu(OMe)] ₂ -OMe	5.85		7.81 (7.1)	7.45 (4.7)							
Boc[Glu(OMe)]Lac[Glu(OMe)] ₃ -OMe			7.93 (7.5)	7.53 (5.0)	7.41 (3.1)						
Boc[Glu(OMe)]Lac[Glu(OMe)] ₄ -OMe	5.80		8.03 (6.9)	7.55 (4.0)	7.55 (4.7)	7.46 (2.9)					
Boc[Glu(OMe)]Lac[Glu(OMe)] ₅ -OMe	5.66*** (4.8)		8.03 (5.9)	7.61 (4.1)	7.55 (2.7)	7.55 (4.3)	7.51 (3.4)				
Boc[Glu(OMe)] ₂ Lac[Glu(OMe)]-OMe	5.84	7.71 (6.6)		7.63 (5.8)							
Boc[Glu(OMe)] ₂ Lac[Glu(OMe)] ₂ -OMe	5.65	7.78 (6.7)		7.70 (5.2)	7.53 (5.6)						
Boc[Glu(OMe)] ₂ Lac[Glu(OMe)] ₃ -OMe	5.70	7.96 (10.2)		7.86 (7.1)	7.69 (5.1)	7.57 (5.0)					
Boc[Glu(OMe)] ₂ Lac[Glu(OMe)] ₄ -OMe	5.73	7.87 (6.3)		7.84 (5.9)	7.80 (8.5)	7.56 (4.2)	7.44 (1.4)				
Boc[Glu(OMe)] ₂ Lac[Glu(OMe)] ₅ -OMe	6.10 (6.3)	8.04 (6.2)		7.89 (1.6)	7.85 (6.6)	7.70 (2.8)	7.61 (3.4)	7.37 (-0.4)			

(Continued)

Table 3. (Continued)

Depsipeptide	Chemical shifts and temperature coefficients*										
	Glu 1**	Glu 2	Glu 3	Glu 4	Glu 5	Glu 6	Glu 7	Glu 8	Glu 9	Glu 10	Glu 11
Boc[Glu(OMe)] ₃ Lac[Glu(OMe)]-OMe	5.75	7.85 (7.3)	7.72 (6.6)		7.68 (6.4)						
Boc[Glu(OMe)] ₃ Lac[Glu(OMe)] ₂ -OMe	5.85	7.95 (7.7)	7.78 (5.9)		7.75 (6.7)	7.52 (5.3)					
Boc[Glu(OMe)] ₃ Lac[Glu(OMe)] ₃ -OMe	5.84	8.03 (7.8)	7.91 (6.7)		7.80 (6.8)	7.49 (4.2)	7.48 (4.0)				
Boc[Glu(OMe)] ₃ Lac[Glu(OMe)] ₄ -OMe		8.09 (7.8)	8.07 (7.2)		7.85 (7.1)	7.70 (6.5)	7.56 (4.5)	7.43 (2.1)			
Boc[Glu(OMe)] ₃ Lac[Glu(OMe)] ₅ -OMe	6.05 (7.1)	8.24 (7.9)	8.14 (7.1)		7.96 (7.8)	7.91 (7.0)	7.72 (4.6)	7.63 (9.1)	7.37 (1.0)		
Boc[Glu(OMe)] ₄ Lac[Glu(OMe)] ₄ -OMe	6.14	8.08 (7.5)	8.00 (7.4)	7.93 (5.6)		7.99 (9.6)	7.74 (7.0)	7.56 (4.3)	7.43 (2.7)		
Boc[Glu(OMe)] ₄ Lac[Glu(OMe)] ₅ -OMe	6.14 (7.8)	8.20 (7.5)	8.07 (7.4)	8.00 (5.9)		8.00 (7.9)	7.93 (7.5)	7.73 (4.8)	7.63 (1.2)	7.36 (7.8)	
Boc[Glu(OMe)] ₅ Lac[Glu(OMe)] ₅ -OMe	6.34 (9.4)	8.27 (6.5)	8.20 (8.6)	8.13 (8.0)	7.99 (5.7)		7.86 (3.7)	7.86 (3.7)	7.78 (4.6)	7.66 (4.5)	7.36 (1.0)

*The values in the second line are temperature coefficients (ppb/K).
**The resonances in this column arise from the NH of a urethane linkage.
***The resonance at higher temperature.

assigned to Glu 4, Glu 5 and Glu 6 respectively, have a small upfield shift compared to oligopeptides.

In the series Boc-[Glu(OMe)]₂-Lac-[Glu(OMe)]_nOMe ($n = 1$ to 5) the urethane amide resonance can be assigned to the highest field resonance in the amide region. In the tetradepsipeptide ($n = 1$) the other two resonances at 7.71 and 7.63 ppm can be assigned to Glu 2 and Glu 4 amide protons in comparison with POE bound oligopeptide studies. The amide proton for the second residue, which precedes lactic acid in this series, is found each time at a higher field than the same proton in the analogous oligopeptide. Analogous to the oligopeptides, the Glu 2 NH resonances in the depsipeptides shift to a lower field as the chain length increases. The resonances appear at 7.78, 7.96, 7.87 and 7.82 ppm respectively.

In the same series, i.e. Boc-[Glu(OMe)]₂-Lac-[Glu(OMe)]_n-OMe ($n = 1$ to 5), the resonances of Glu 4 which follow lactic acid are shifted in the opposite direction in comparison to the equivalent proton in the analogous oligopeptides. The inductive effect of the lactic acid dies out after the second residue away from the lactic acid residue. In fact, as one considers the residues further toward the C-terminal end, a high field shift of resonances is observed as compared to analogous oligopeptides. On the amino terminal side, the amide proton adjacent to the lactic acid residue appears at 7.70, 7.86, 7.84 and 7.63 ppm in the depsipeptides series when $n = 2$ to 5. These resonances for Glu 2 NH appear at 7.67, 7.75, 7.76 and 7.86 ppm in the analogous oligopeptides. Thus, two opposing effects are operative. As the chain length increases in the depsipeptides, the inductive effect of the lactic acid residue shifts the Glu 2 NH to a higher field. In the peptide series as the chain length and order increases, the Glu 2 NH shifts slightly to a lower field. These effects can be understood only after the assessment of secondary structure by circular dichroism (see CD Section). The C-terminal amide protons in the depsipeptides are also high field shifted when compared to the respective oligopeptides. The origin of these differences is not well understood.

In the series Boc-[Glu(OMe)]₃-Lac-[Glu(OMe)]_n-OMe ($n = 1$ to 5) the two low field resonances were assigned to Glu 2 and Glu 3 amide protons. The Glu 3 residue precedes the lactic acid residue in this family. Thus, the amide resonance of this residue has low field shift as compared to the analogous oligopeptide molecules. Similar results were seen for the Glu 2 residue in the previous series where Glu 2 is the residue preceding the lactic acid. The residue that succeeds the lactic acid (i.e. Glu 5) experiences an inductive effect which leads to a low field shift of this resonance compared to the analogous oligopeptide. The assignments of the other resonances in this series were carried out using a similar approach as described above.

Temperature coefficients of various amide proton resonances for the series of depsipeptides in trifluoroethanol are also presented in table 3. The urethane amide proton (where measurements were possible) has a temperature coefficient in the range of 7 to 9 ppb/K. This proton, as in the oligopeptide series, seems to be solvent exposed. The temperature coefficients in depsipeptides may also be divided into two categories: the N-terminal protons which have large values and the C-terminal protons which have small values. The temperature coefficients of the C-terminal amide protons become smaller as the chain length is increased. The oligopeptide chain seems to be necessary for the small temperature coefficients indicative of the secondary structure of the molecule. In the undcapeptide the temperature coefficients of all the amides except the urethane are in the range of -0.1 to 6.2 ppb/K whereas in the

undecadepsipeptide the N-terminal amides are within the range of 5.7 to 9.4 ppb/K and the C-terminal amide protons are in the range of 1.0 to 4.6 ppb/K. The effect of the lactic acid residue in the center of the undecadepsipeptide is to make the N-terminal residues more solvent exposed. The secondary structure is maintained, as shown in the circular dichroism section below.

Circular Dichroism

Circular dichroism is an excellent tool for studying the secondary structures of peptide molecules (12, 13) while NMR methods give information about structure at the residue level and the nature of hydrogen bonds of amide protons. End groups affect the appearance of the CD spectra more significantly in the short peptides (12) and much less so in the larger polypeptides (14). Trifluoroethanol is known to support α -helical and β -sheet structures for oligopeptides (12, 14). The earlier studies of γ -ethyl-L-glutamates (15) and L-methionine (16) show that these peptides attain α -helical structures at about heptamer chain lengths in TFE. The oligopeptides of alanine, isoleucine and valine have been shown to attain β -structures in TFE (17–19).

We report in this paper the CD studies of oligoglutamates and depsipeptides in TFE and compare these results with NMR studies of these peptides in the same solvent. These studies have been carried out at an approximate concentration of 10^{-3} mol/l. Figure 2 shows the CD spectra of oligo- γ -methyl glutamates

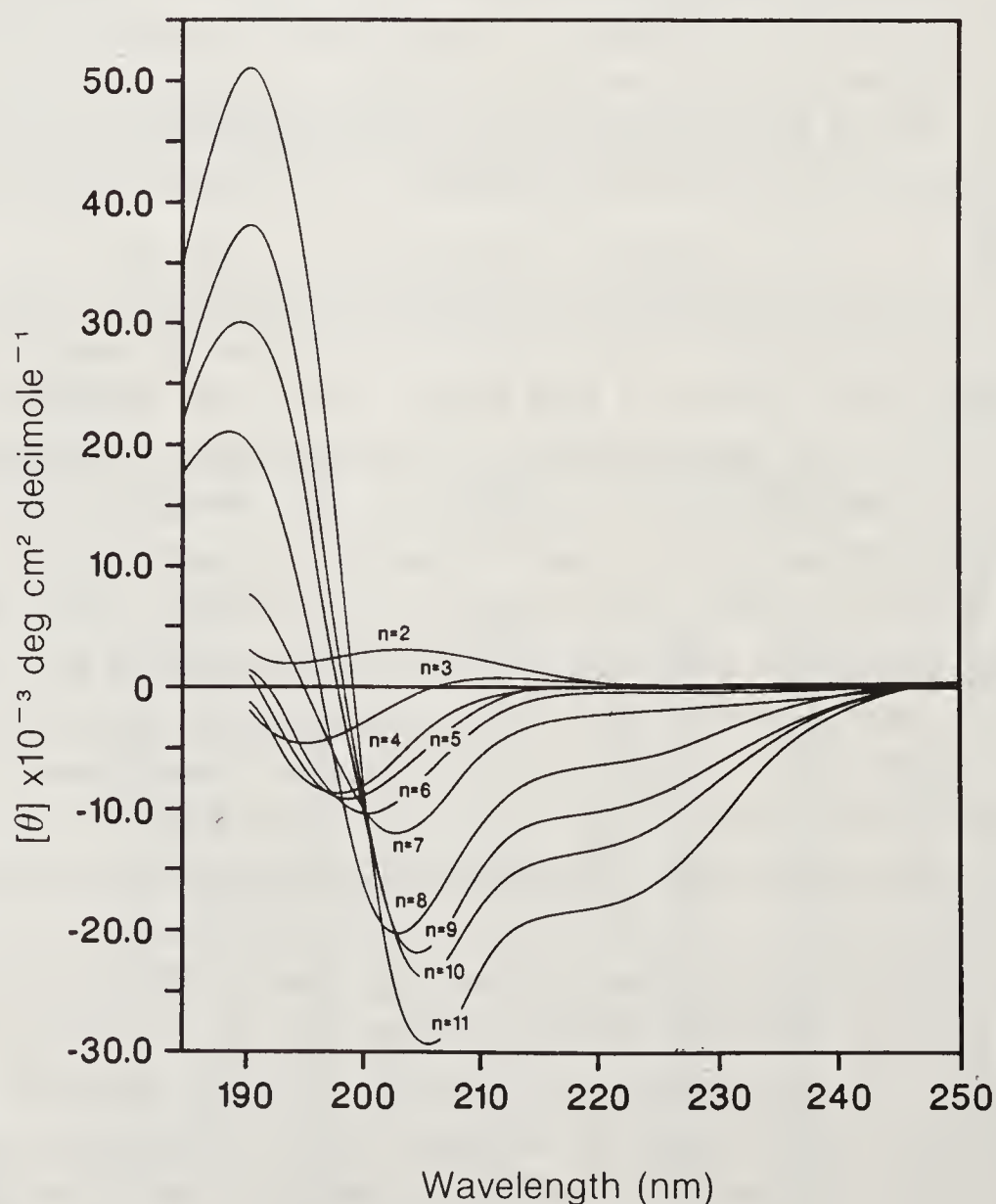


Figure 2. CD spectra of Boc-[Glu(OMe)]_nOMe ($n = 2$ to 11) in trifluoroethanol. The concentrations are around 1 millimolar.

Boc-[Glu(OMe)]_n-OMe ($n = 2$ to 11). In the smaller peptides, $n = 2, 3, 4, 5$ and 6, no ordered structure was observed. In these peptides there is basically one negative band near 200 nm which is typical for non-ordered structures (20). The onset of helicity can be seen from the heptamer onwards. The negative band arising from the excitation of the $\pi - \pi^*$ transition shifts to 205 nm from 198 nm (in smaller peptides). Poly-lysine in the α -helical conformation has this band at 208 nm (20). Based on the ellipticity at 208 nm, the undecaglutamate has about 84% helicity (20). The negative band at 220 nm is enhanced in intensity commencing with the heptamer. As the chain length grows, the band arising from the $n \rightarrow \pi$ transition continues to grow.

The CD spectra of the series of oligodepsiglutamates are shown in figure 3. The first series (figure 3a) contains Boc-[Glu(OMe)]-Lac-[Glu(OMe)]_n-OMe ($n = 1$ to 5). As noticed in our earlier studies (10) there is a regular increase in intensity of the

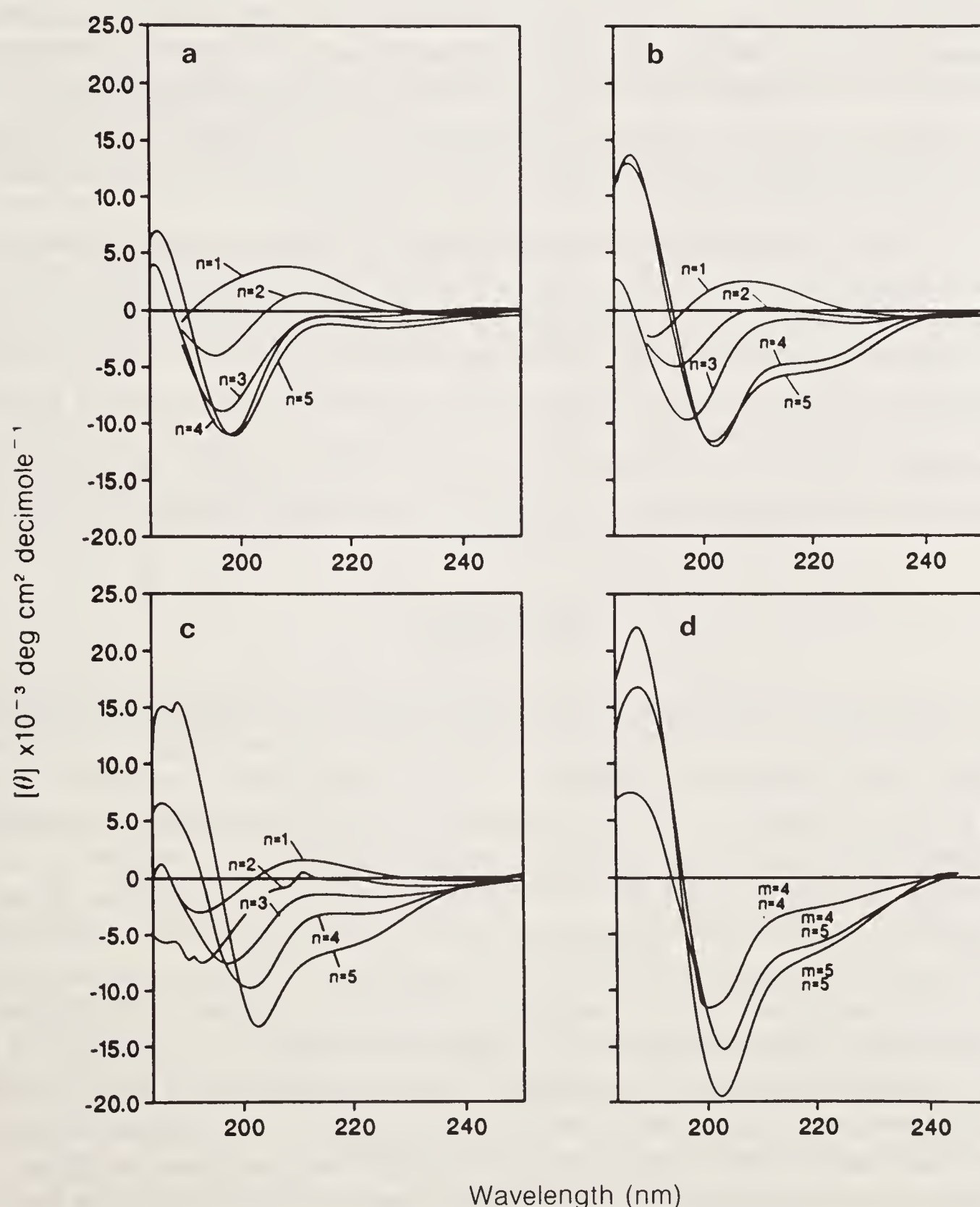


Figure 3. CD spectra of Boc-[Glu(OMe)]_m-Lac-[Glu(OMe)]_n-OMe ($m = 1$ to 5 and $n = 1$ to 5) in trifluoroethanol. The concentrations are around 1 millimolar.

negative band near 200 nm. A band near 197 nm is indicative of random coil structures in peptides (20). In the tripeptide ($n = 1$) the positive band around 210 nm is mainly due to $n - \pi^*$ transition of the ester bond. With an increase in chain length, the effect of the ester bond decreases and the $\pi - \pi^*$ transition of the amide becomes the main contributor of this band. No ordered structures can be attributed in the first series of peptides.

In the series Boc-[Glu(OMe)]₂-Lac[Glu(OMe)]_n-OMe ($n = 1$ to 3) (figure 3b) a negative band is observed below 200 nm indicative of disordered structures. At the heptapeptide and octapeptide ($n = 4$ and 5) stage this band shifts to 202 nm showing the development of ordered structure. A shoulder around 220 nm was also seen in the case of the octapeptide. The intensities of the negative bands in the CD spectra of depsipeptides were greatly affected by the positive band of the $n - \pi^*$ transition at 210 nm generated by the ester bond.

The penta, hexa and heptamer members of the series Boc-[Glu(OMe)]₃-Lac[Glu(OMe)]_n-OMe ($n = 1$ to 5) also have random coil structures as indicated by the position of the negative band below 200 nm (figure 3c). The $\pi - \pi^*$ band of the octapeptide and nonapeptide ($n = 4$ and 5) shifts to 202 nm and the negative band close to 220 nm was also observed indicating the development of the α -helical structures. Once again, the presence of the ester band makes it very difficult to calculate the percentages of helicity for these depsipeptides.

In the longer depsipeptides, Boc-[Glu(OMe)]_m-Lac-[Glu(OMe)]_n-OMe where $m = 4$ or 5 and $n = 5$, the $\pi - \pi^*$ appears at 202 nm and the $n - \pi^*$ transition of the amide was clearly observable around 220 nm. For these molecules, a strong positive band at 189 nm was also present. This positive band is part of the exciton splitting of the 190 nm $\pi - \pi^*$ transition (21, 22). This band has been observed in spectra of peptides with α -helical structures while it was absent in spectra of random coil structures (20). Thus, we can state that the presence of the lactic acid in the middle of these longer depsipeptides does not destroy the α -helical secondary structure.

Conclusion

These results show that in oligo- γ -alkyl glutamates, as the chain length increases, the presence of ordered structure increases in solution similar to that of poly- γ -alkyl glutamates. The temperature coefficients of the amide proton resonances become smaller and the ellipticities at 222 nm increase. The ¹H-NMR results are presumably due to intramolecular hydrogen bonding which is indicative of ordered structures. The circular dichroism results for the oligopeptides are also indicative of ordered structures but, of course, are determined by the overall molecular conformations. Helicity leads to the increase of the $n \rightarrow \pi^*$ transitions (~ 222 nm) and the splitting of the $\pi \rightarrow \pi^*$ transitions with a concurrent red shift (~ 206 nm). Introduction of a lactic acid residue in the structures of the oligopeptides disrupts the ability of a given residue to donate hydrogen in the α -helical structures. However, it does not appear that incorporation of lactic acid destroys the α -helix in the longer depsipeptide studied where more stable α -helices exist. The ellipticities are lower at 222 nm. This is partially due to the positive band of the ester bond around 210 nm. The lactic acid residue does destabilize helices of shorter peptides which are just at or above the critical size necessary for helicity.

Acknowledgements

This work was supported by NIH grant GM 18694 and NSF grant DMR 8816326. This paper is dedicated to Prof. G. N. Ramachandran.

References

1. Becketl, W. J., Mathias, L. J. & Goodman, M. (1980) *Macromolecules* **14**, 203–207.
2. Goodman, M., Becketl, W. J., Katakai, R. & Wouters, G. (1981) *Macromolecules Chem. Suppl.* **4**, 100–115.
3. Katakai, R. & Goodman, M. (1982) *Macromolecules* **15**, 25–30.
4. Ribeiro, A. A., Saltman, R. & Goodman, M. (1985) *Biopolymers* **15**, 2495–2510.
5. Neises, B. & Steglich, W. (1978) *Angew. Chem. Int. Ed. Engl.* **17**, 522–524.
6. Hore, P. (1983), *J. Magn. Reson.* **55**, 283–300.
7. Ribeiro, A. A., Goodman, M. & Naider, F. (1978) *Int. J. Pept. Protein Res.* **14**, 414–436.
8. Naider, F., Sipzner, R., Steinfeld, A. S., Becker, J. M., Ribeiro, A. A. & Goodman, M. (1979) in *Proceedings of 6th American Peptide Symposium*, eds. Gross, E. & Meinenhofer, J. (Pierce Chemical Co., Rockford), pp. 185–188.
9. Ribeiro, A. A., Saltman, R. P. & Goodman, M. (1980) *Biopolymers* **19**, 1771–1790.
10. Goodman, M., Venkatachalapathi, Y. V., Mammi, S. & Katakai, R. (1985) *J. Biosci.* **8**, 223–238.
11. Pysh, E. S. & Toniolo, C. (1977) *J. Am. Chem. Soc.* **99**, 6211–6219.
12. Naider, F. & Goodman, M. (1977) in *Biorg. Chem: Macro- and multi-molecular systems*, ed. Van Tamelen, E. E. (Academic Press, New York), Vol. 3, pp. 177–202.
13. Woody, R. W. (1985) in *The Peptides*, ed. Hruby, V. J. (Academic Press, New York), **7**, pp. 15–114.
14. Bode, K., Mutter, M., Saltman, R., Goodman, M. & Riberio, S. (1983) *Biopolymers* **22**, 163–169.
15. Goodman, M., Verdinni, A. S., Toniolo, C., Phillips, W. D. & Bovey, A. (1969) *Proc. Natl. Acad. Sci. U.S.A.* **64**, 444–450.
16. Bonora, G. M. & Toniolo, C. (1974) *Biopolymers* **13**, 2179–2190.
17. Goodman, M., Rupp, R. & Naider, F. (1971) *Bioorg. Chem.* **1**, 294–309.
18. Goodman, M., Naider, F. & Toniolo, C. (1971) *Biopolymers* **10**, 1719–1730.
19. Toniolo, C., Bonora, G. M. & Fontana, A. (1974) *Int. J. Pept. Protein Res.* **6**, 371–380.
20. Greenfield, N. & Fasman, G. D. (1969) *Biochemistry* **8**, 4108–4116.
21. Moffit, W. (1956) *J. Chem. Phys.* **25**, 467–478.
22. Tinoco, I. Jr., Woody, R. W. & Bradley, D. F. (1963) *J. Chem. Phys.* **38**, 1317–1325.

Peptides as Models for Protein-Folding

E. Subramanian

Department of Crystallography and Biophysics,
University of Madras, Guindy Campus, Madras 600 025, India

Introduction

The problem of how proteins fold up to acquire their unique three-dimensional structure has been the subject of intensive research for the past few decades. Anfinsen (1) showed that the information about the folding is contained in the sequence of amino acids constituting the protein chain. The experiments also implied that the folding process may be seeded – that is, parts of a protein chain can act as nucleation sites for folding. Such nucleation sites are likely to be characterized by short peptide sequences possessing dominant conformational states. The implication is that even an unfolded protein chain may have significant ordered structure. If one can identify folding intermediates and recognise the peptide fragments constituting nucleation sites, analysis of their structures may tell us which interactions stabilize these structures at the beginning of the folding pathway.

The implications of structural studies on peptide fragments extend beyond understanding the mechanism of protein folding. Structural studies on peptides in general are useful for theoretical modelling studies designed to predict the three-dimensional structure of a protein purely from its amino acid sequence. The theoretical studies generally involve a statistical analysis of the results of X-ray crystallographic studies on proteins in order to develop predictive algorithms for correlating certain amino acid sequences to certain structural features. Presently, X-ray structures are available for more than 300 proteins (2); however many of these studies involve homologous proteins, hence a satisfactory development of the rules for correlating amino acid sequence to structural features remains a distant objective, pointing to the need for expansion of the data-base of known protein structures. Protein crystal structure analysis is however a time-consuming affair. In the interim, thanks to recombinant DNA technology, the data-base of known protein sequences is growing at an exponential rate, while the acquisition of structural knowledge through protein crystallography moves at a snail's pace. Moreover, biochemists have started to compile a catalogue of all the genes in the human genome. The gene sequences translate into amino acid sequences and hence to various proteins. Only by knowing how a given sequence of amino acids might fold, can one predict the biological function of the corresponding protein.

There is thus an urgent need for developing accurate prediction rules to correlate sequence to structure. Theorists, therefore, find it expedient to use the results of X-ray

studies on small peptides as models for deducing stable conformations of proteins. The validity of this approach stems from the classical work of Pauling and his associates (3) who predicted the right-handed alpha helix and the pleated beta sheet structure as two possible stable conformations for the peptide backbone of a protein chain.

Structural studies on peptides are important for other reasons too. Peptides play important biological roles such as hormones, antibiotics, toxins, neurotransmitters, etc. Many small peptides have been found to be modulators of a wide variety of biological functions. These peptides (containing two or more residues) are often the products of processing of much larger precursor proteins. Many biological functions such as digestion, blood pressure regulation, sexual maturation and reproduction, learning and memory, hunger and pain are all now considered to be mediated by peptides. The potency and specificity of peptides in these biological functions make it clear that peptides and their analogues will be among the major drugs of the future. The conformation of the peptide molecules would appear to be the major determinant in the functional properties of the peptides.

From the above discussion it is clear that there is a great deal of interest in developing the structural studies on peptides. The most direct way to establish the structure and conformation of a molecule in an *unambiguous* way is by the methods of X-ray crystallography. In this article, therefore, an attempt has been made to summarize the results of X-ray crystallographic studies on peptides. Excellent review articles on the subject of peptide conformation are already available in the literature (4, 5), and consequently no efforts have been made to provide an exhaustive survey of the literature. Instead, this article will highlight the results of X-ray studies on linear peptides carried out by the author and his associates at Madras.

The essential structural features of a peptide chain are determined by the planar amide units constituting the backbone of the chain. The folding of a peptide chain is accomplished by rotations about $C^\alpha-N$ and $C^\alpha-C'$ bonds at each α -carbon atom, the respective torsion angles being referred to as phi (ϕ) and psi (ψ) (figure 1). Deviations from planarity of the peptide bond is indicated by the torsion angle omega (ω), corresponding to rotation about the peptide bond $C'-N$. The convention for the torsion angles follows the recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature (6). According to this convention, an extended backbone with *trans* planar peptide units has all the torsion angles, phi, psi and omega, equal to 180° . In the following discussion, all amino acids are assumed to be L-enantiomers, unless otherwise indicated. Moreover, the survey is generally restricted to a discussion of the conformational features of the backbone of the peptide chains, and unless relevant, conformational aspects of the side chains will not be considered, since these conformations generally follow predictable lines based on stereochemical requirements and side chain chemical character. It may be noted that a tripeptide, consisting of two linked peptide units, is perhaps the smallest fragment of a peptide chain in which the effects of immediately neighbouring residues on the conformational preferences of a given residue can be investigated, leading to the recognition of the existence of sequence-dependent conformational states.

Since X-ray studies pertain to the solid state, one may wonder if the conformational features deduced may be relevant to bioactive solution conformations of the peptides. Even the analysis of solution conformations is vitiated by conformational heterogeneity and dynamics. Linear peptides are highly flexible molecules that can undergo conformational interconversions. NMR is often the most useful method of conform-

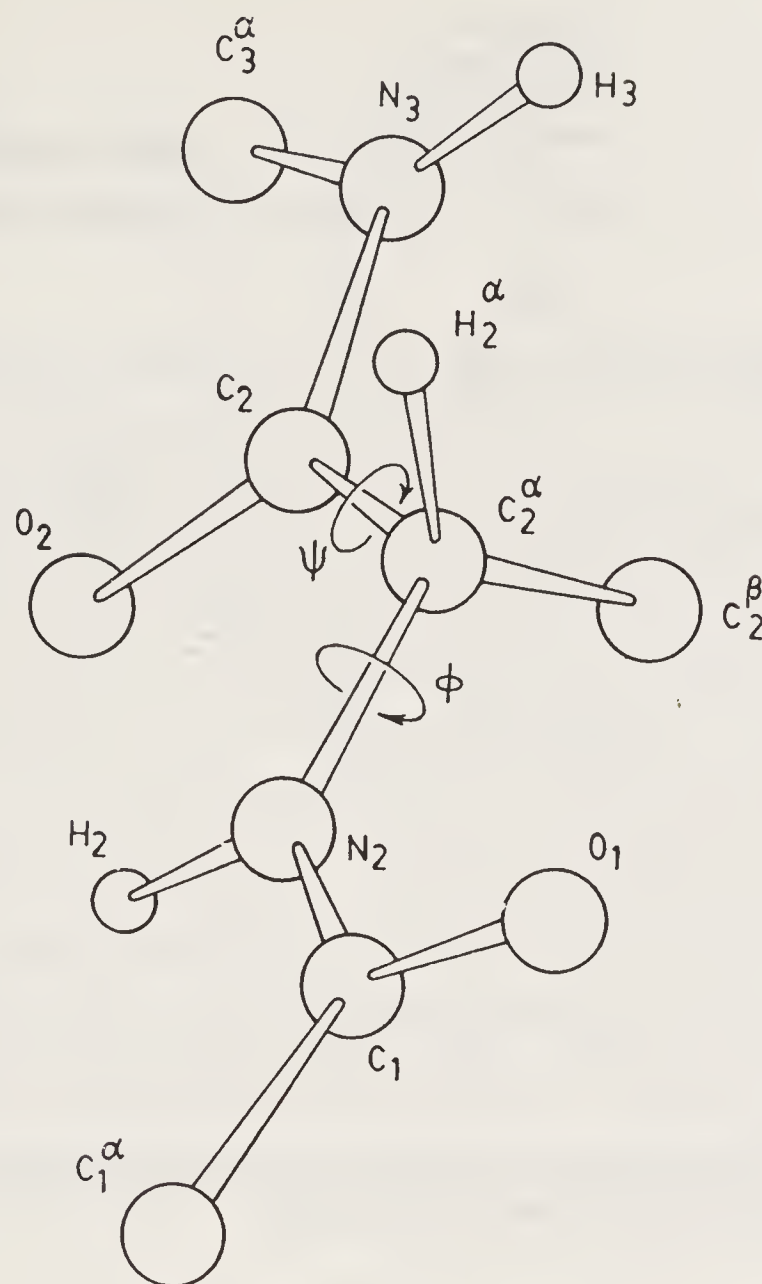


Figure 1. Definition of the torsion angles at each alpha carbon atom.

ational analysis in solution, but with peptides, conformational interconversions not requiring peptide bond rotations could lead to averaged parameters, whose interpretation in terms of a single conformation may have little value (5) unless correlated with X-ray studies. Quite often, the X-ray studies themselves have revealed conformational heterogeneity, as was found in the case of the enkephalins (7–11).

General Conformational Features of Peptide Backbones

Crystallographic studies reveal that the observed phi, psi angles for the backbone of a peptide chain generally occur within the allowed regions of the conformational space as predicted by the Ramachandran diagram (figure 2) (12). In general, the peptide bond geometry is found to be essentially the same from one residue to another, and in different peptides. The peptide unit is generally planar ($\Delta\omega = 0$), but deviations as large as 10 to 15° have been reported. The peptide bond is generally *trans* with occasional exceptions, especially when proline or sarcosine is involved. If the peptide termini are not protected, then the molecule exists as a zwitterion with terminal NH_3^+ and COO^- groups. If the N-terminal alone is protected, then the carboxyl terminal is un-ionized and is characterized by the following features: $\text{C}'\text{—OH}$ bond is more than 0.1 Å longer than $\text{C}'\text{=O}$ bond and the angle $\text{C}^\alpha\text{—C}'\text{—O(H)}$ is nearly 10° smaller than that

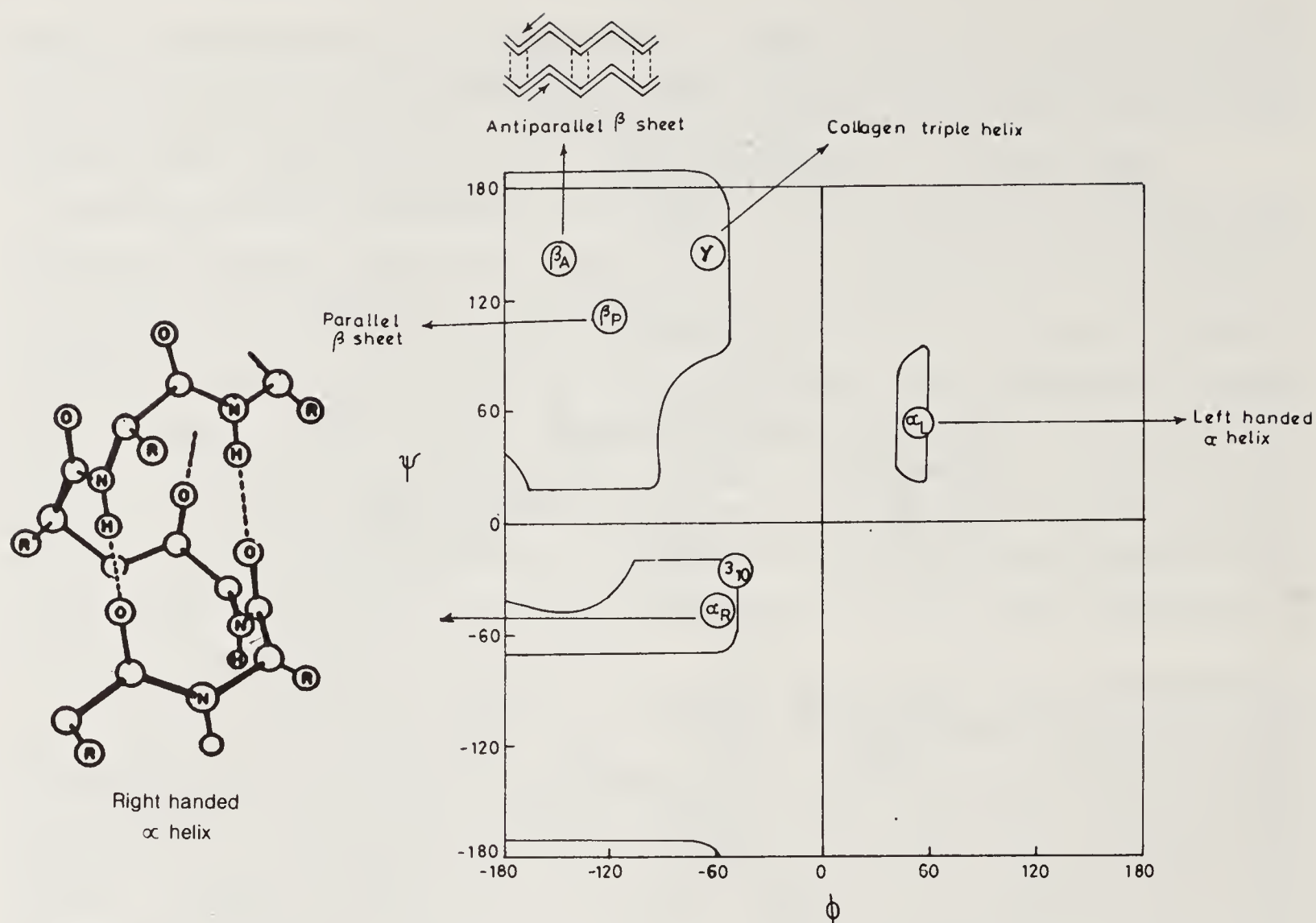


Figure 2. Ramachandran Diagram representing allowed regions of the conformational space for the backbone of a peptide chain.

observed in ionized carboxyl groups. Specifically, the observed backbone conformation in small linear peptides can be classified into three categories:

- a) extended; b) helical; c) folded

A brief description of the observed features in each category is given below:

Extended Conformations

In theory, a fully extended all *trans* peptide backbone has $\phi, \psi = 180^\circ, 180^\circ$. In practice, there is a slight folding at each alpha carbon atom, presumably to facilitate two extended peptide chains to interact with each other laterally through intermolecular hydrogen bonds. This gives the appearance of a pleated sheet to the structure, hence this conformation is referred to as a "beta pleated-sheet structure".

Depending upon the relative directions of the two interacting chains, one can have either parallel beta-sheet or antiparallel beta-sheet structure.

Parallel beta-sheet structures: If two extended peptide chains interact with their chain directions parallel to each other, this gives rise to the parallel beta-structure. If the number of interacting chains are large, one gets an infinitely extended beta-sheet structure. For such structures, Pauling *et al.* (3) found, from model building studies,

that:

phi, psi	= $-119^\circ, 113^\circ$
inter-chain H-bond	= 2.87 \AA
interchain spacing	= 4.85 \AA

The chain repeat distance, defined as the distance between every second alpha-carbon atoms (that is $C_i^\alpha - C_{i+2}^\alpha$) was found to be 6.5 \AA .

The first example of a parallel beta structure in linear peptides, in close agreement with Pauling's proposal, was that observed in glycyl-L-phenylalanyl-glycine by Marsh and Glusker in 1962 (13), with $(\phi, \psi) = (-126^\circ, 132^\circ)$, a chain repeat distance of 6.74 \AA , interchain spacing = 4.90 \AA and two hydrogen-bonds of lengths 2.89 \AA and 3.00 \AA . The most extended parallel beta structure has been observed in the crystal structure of L-valyl-glycyl-glycine (14), with $(\phi, \psi) = (-155^\circ, 155^\circ)$, chain repeat distance of 7.23 \AA , a hydrogen bond distance of 3.00 \AA and an inter-chain spacing of 4.85 \AA . The tripeptide, L-isoleucyl-glycyl-glycine, is isomorphous with L-valyl-glycyl-glycine (as judged by crystallographic parameters) and is therefore expected to adopt the same highly extended parallel beta structure (15). In most of the peptide structures displaying parallel beta sheet arrangement, the interchain distance of 4.85 \AA frequently corresponds to one of the lattice-repeat distances. In other words, interchain hydrogen bonds occur between adjacent peptide molecules related by a unit cell translation, giving rise to a parallel beta-sheet of infinite extension. Figure 3 shows a stereo view of the parallel beta sheet structure observed in the crystal of L-valyl-glycyl-glycine (14).

The structure of L-valyl-glycyl-glycine described above was used as a model for parallel beta sheet structure, by Bandeker and Krimm (16) in their efforts to verify their theoretical predictions of the normal mode spectrum for parallel chain beta sheet structure. These authors showed that normal mode analysis can be used to interpret IR and Raman spectra of peptides and proteins in terms of the peptide chain conformation. They have used such analyses to demonstrate the presence of standard secondary structural features such as alpha helix, anti-parallel beta sheet, beta turns, etc. Until recently, their analysis was not applied to deduce the presence of parallel-chain beta sheet structure, since this structure had not as yet been identified in synthetic polypeptides. Using known structural parameters of the parallel-chain beta sheet arrangement of valyl-glycyl-glycine, Krimm and his associates (16) have now been able to place their own analysis in perspective. The agreement between predicted and observed frequencies and intensities in the IR and Raman spectrum of the tripeptide have led these authors to claim enough confidence in the application of force fields to the calculation of the vibrational modes of the general parallel-chain beta-sheet structure.

Anti-parallel beta structures: From fibre diffraction studies, Arnott *et al.* (17) proposed the following parameters for antiparallel beta chains in peptides:

phi, psi	= $-139^\circ, 135^\circ$
chain repeat	= 6.95 \AA
interchain H-bond	= 2.77 \AA

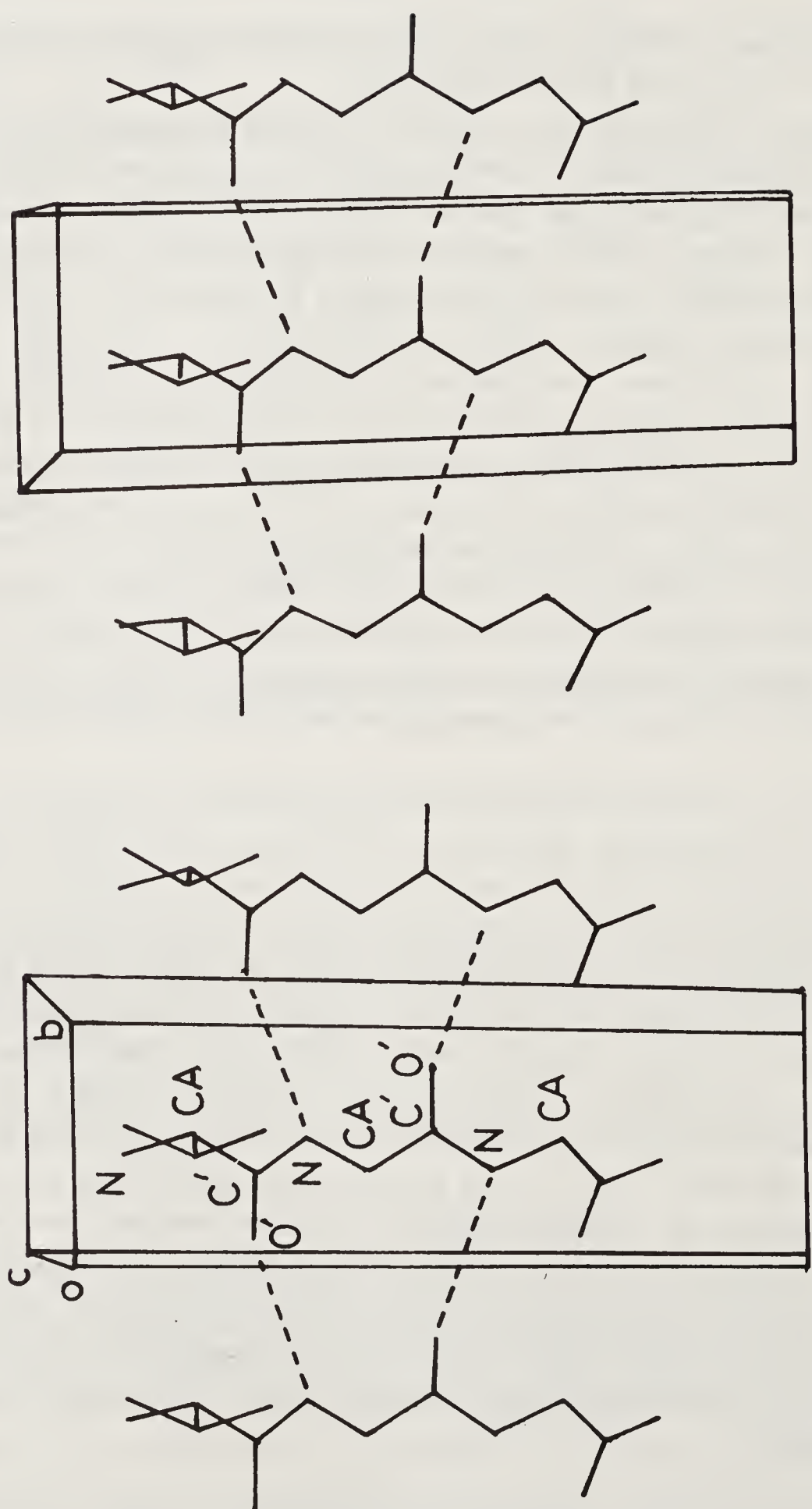


Figure 3. Stereoscopic view of the parallel beta structure observed in the crystal structure of L-valyl-glycyl-glycine (14).

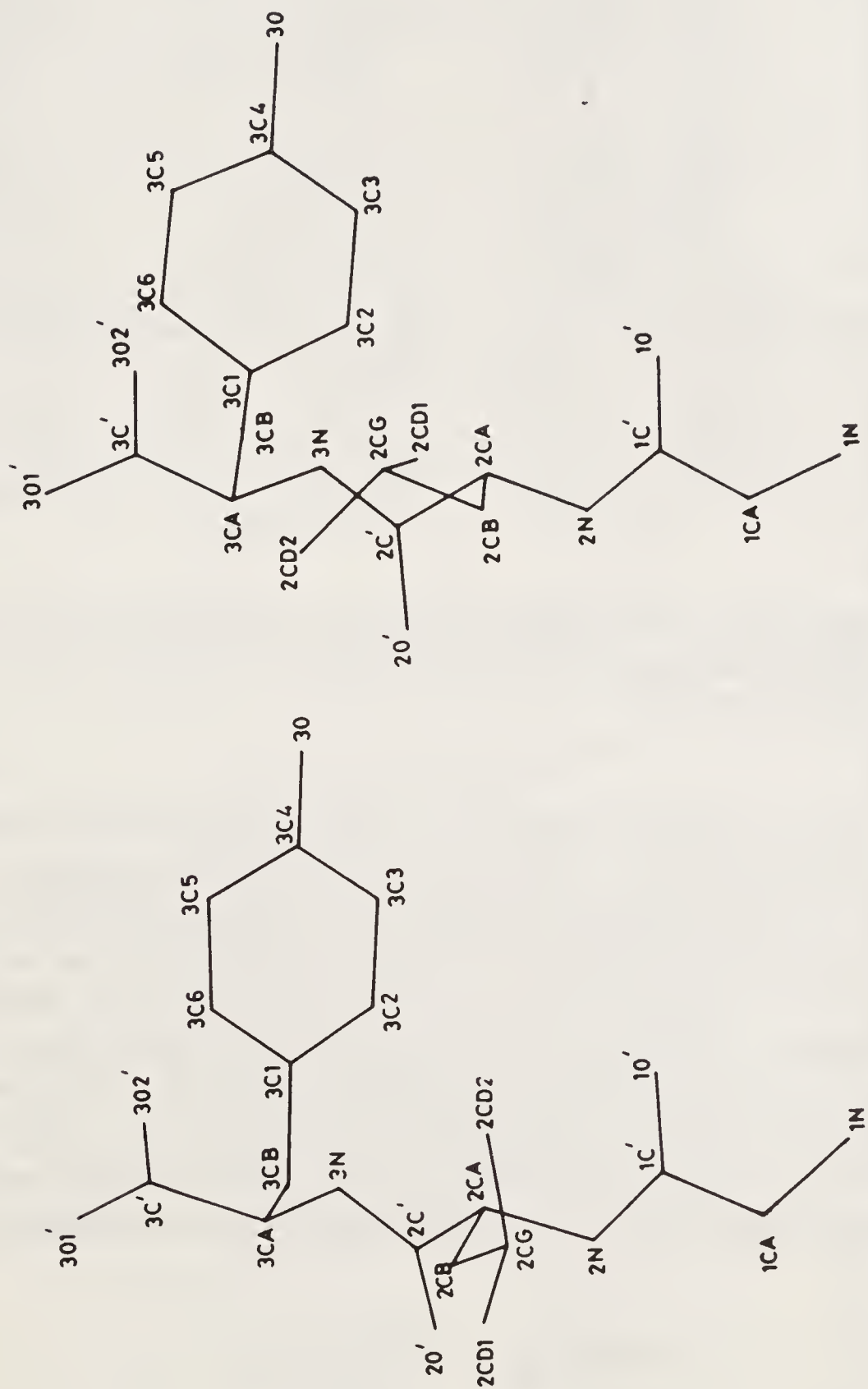


Figure 4. Observed conformations for the two crystallographically independent molecules of glyceryl-L-leucyl-L-tyrosine (18).

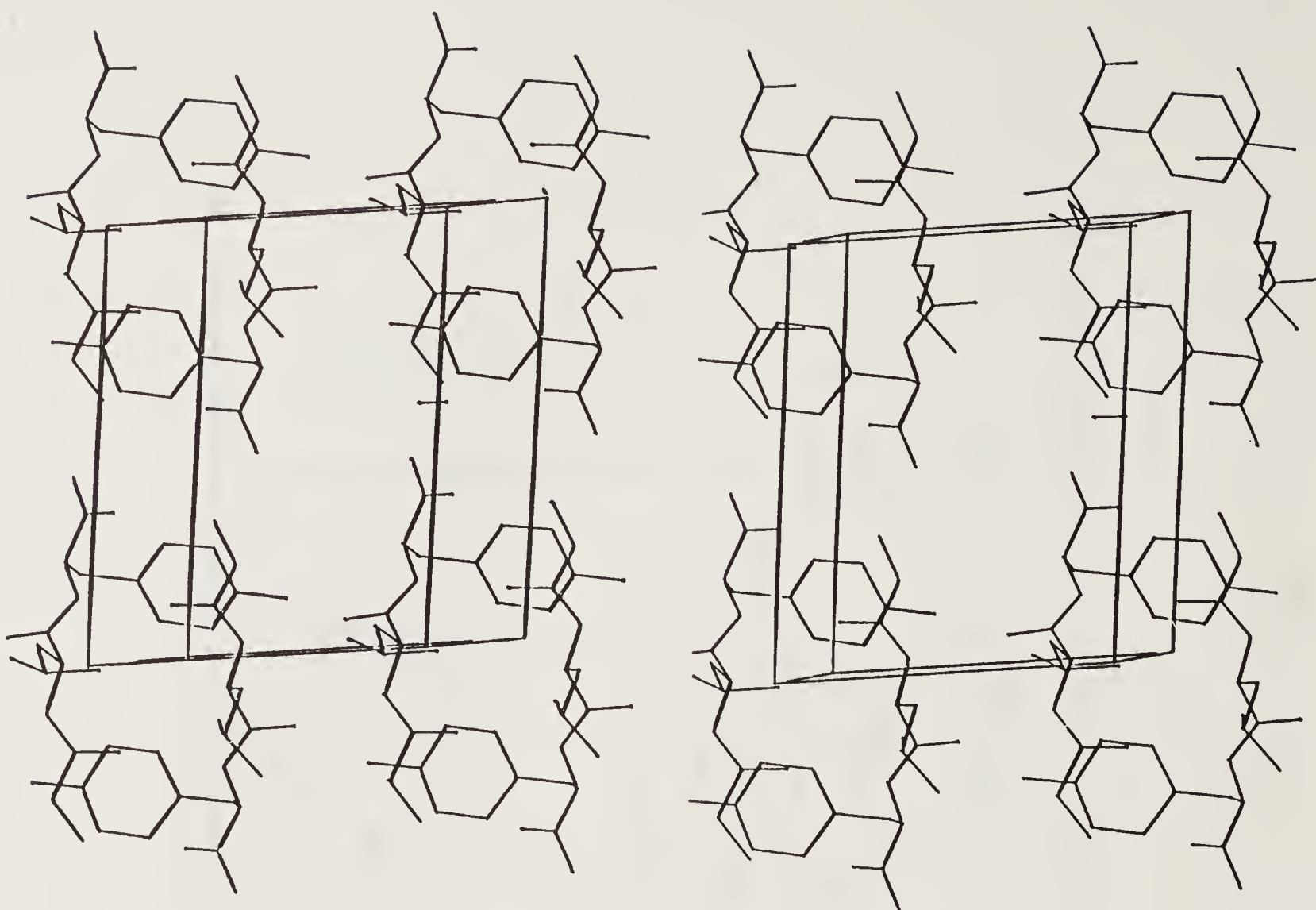


Figure 5. Stereoscopic view of molecular packing in the crystal structure of glycyl-leucyl-tyrosine (18), to illustrate the lateral association of two peptide chains to form an antiparallel beta structure.

The crystal structure of glycyl-L-leucyl-L-tyrosine (18) offers a very good example of antiparallel beta structure. There are two crystallographically independent molecules A and B in a triclinic cell (symmetry P1). The two molecules have similar backbone conformations, but differ mainly in the sidechain conformation at the leucyl residue, as shown in figure 4. The two molecules have their chain directions antiparallel, and since the lattice is triclinic, the crystal packing (figure 5) gives rise to an infinite sheet of antiparallel beta chains of the type ABAB.... with lateral H-bonds along the horizontal direction and head-to-tail interactions along the vertical direction. Figure 6 is a stereoscopic view of the crystal packing viewed along the direction of the hydrogen bonds in the sheet plane. The conformational parameters observed are

	Molecule A	Molecule B
phi, psi	= -131° , 148°	= -146° , 134°
chain repeat	= 6.94	6.88
NH...O	= 2.86, 3.04	3.01, 3.09

This study provides the first direct experimental measurement of the conformational parameters for an antiparallel pleated sheet arrangement of peptide chains possessing bulky side-chains.

For most peptides exhibiting antiparallel beta structure, a two-fold (real or pseudo) rotation axis perpendicular to the chains relates two neighbouring strands in an

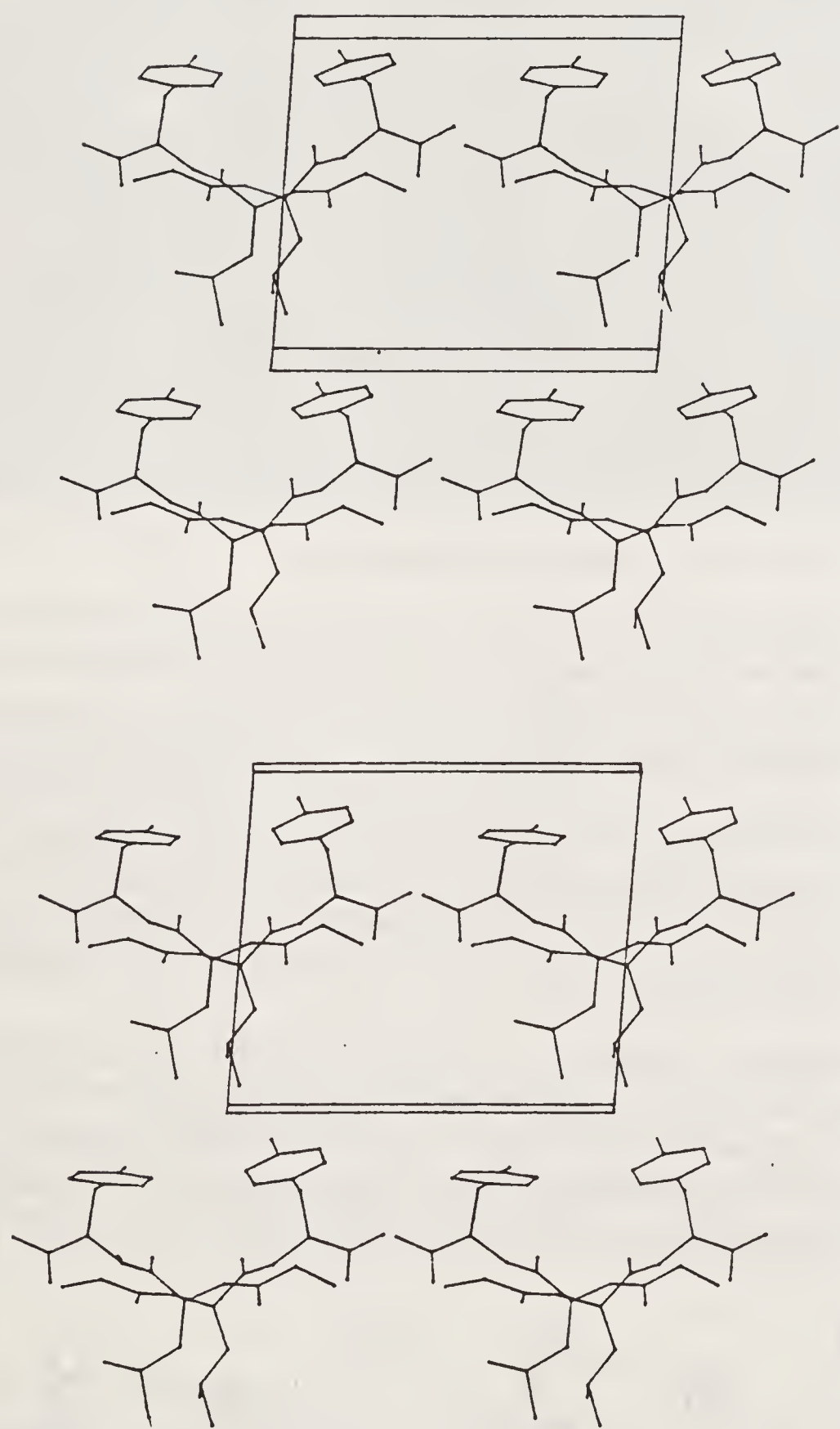


Figure 6. Stereoscopic view of the antiparallel beta structure of glycyl-leucyl-tyrosine (18), viewed in the direction of the hydrogen bonds in the sheet plane.

Table 1. A listing of the backbone conformational angles phi and psi in some tripeptides showing antiparallel beta structure.

Peptide	Molecule A	Molecule B	Reference
Gly-Gly-Gly	$-178^\circ, 172^\circ$	$-166^\circ, 175^\circ$	20
Ala-Ala-Ala	$-145, 145$	$-156, 149$	21
Gly-Leu-Tyr	$-131, 148$	$-146, 134$	18

antiparallel fashion. If all the residues are glycine, such antiparallel association can also be produced by a center of inversion, as observed in the crystal structures of glycyl-glycyl-glycine HCl (19), and glycyl-glycyl-glycine (20).

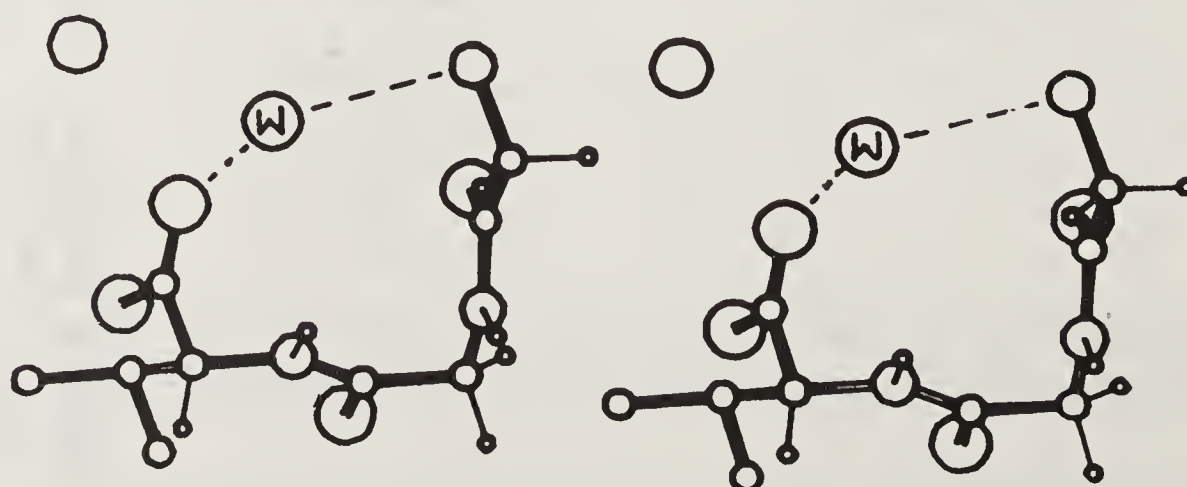
Table 1 shows that the peptide backbone assumes a very extended conformation when the side chains are small and becomes more pleated as the side chains become bulkier. In fact, in the crystal structure of glycyl-glycyl-sarcosine (22), the values of phi, psi at the middle glycine residue are ($-179^\circ, -171^\circ$), even though the molecule does not form any kind of beta structure.

Helical Conformation

Neither the α -helix (with $\phi, \psi = -57^\circ, -47^\circ$), nor the 3_{10} helix (with $\phi, \psi = -60^\circ, -30^\circ$) has been observed in oligopeptides consisting entirely of the naturally occurring amino acids in their sequence at atomic resolution. Both these helical conformations are right-handed in their stable forms and are characterised by internal hydrogen bonds of the type $N-H \cdots O=C$. However, peptides containing α -aminoisobutyric acid (Aib) residues or glycine frequently adopt such helical conformations (23).

In the crystal structure of glycyl-glycyl-L-valine (24), we found that the backbone folding (with $\phi, \psi = -77^\circ, -32^\circ$), is representative of a right-handed 3_{10} -helical fragment. The conformation appears to be stabilised by an intramolecular water bridge (figure 7).

A third type of helical conformation, referred to as the γ -helix by Ramachandran (25) has also been observed in the crystal structures of oligopeptides. The nomenclature was suggested by Ramachandran to refer to a class of regular helical conformations indicated by the diffraction patterns from collagen, polyproline II, polyglycine II as well as from synthetic polymers with repeating sequences such as Gly-Pro-Pro and

**Figure 7.** The molecular conformation of glycyl-glycyl-L-valine (24).

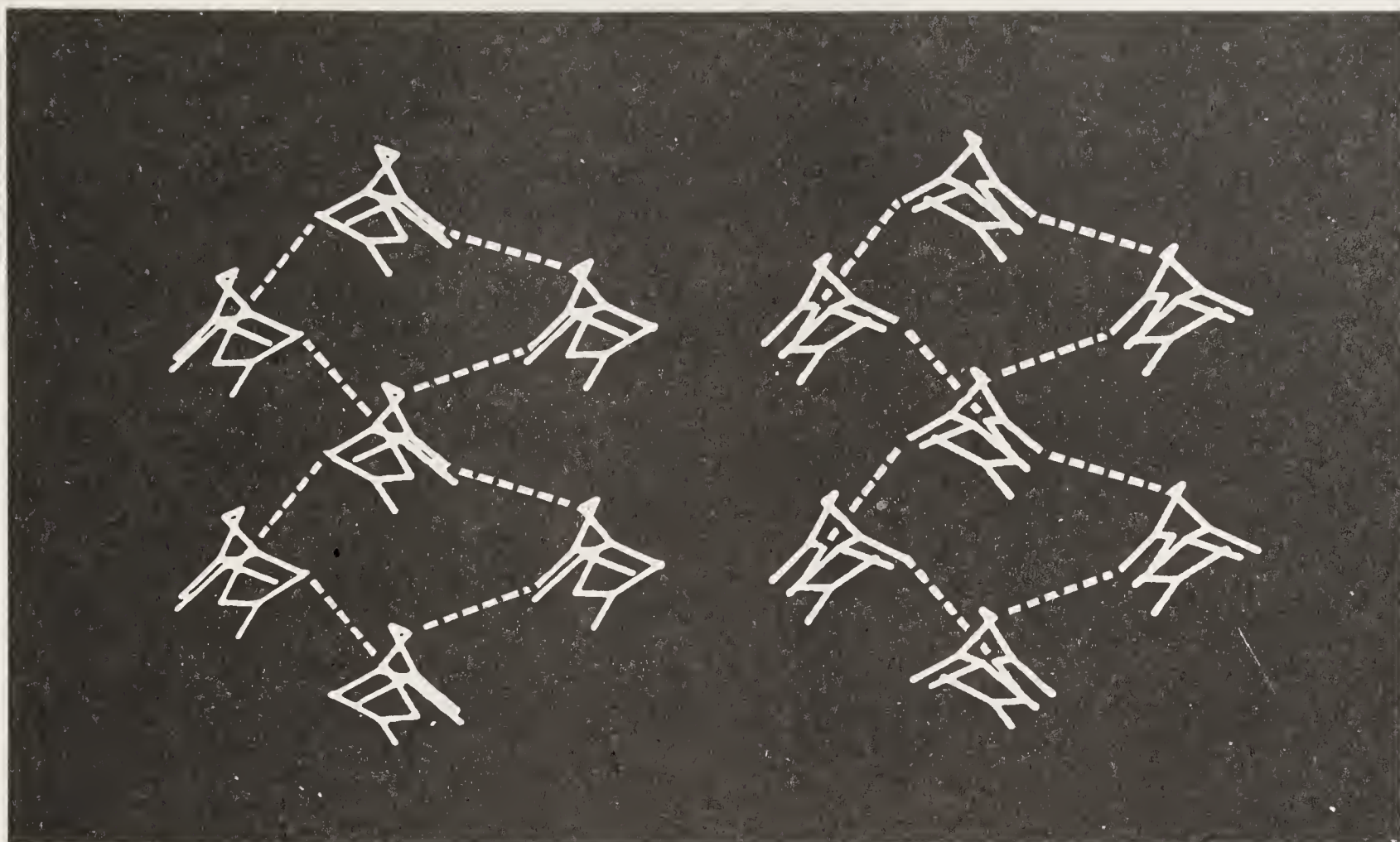


Figure 8. Crystal packing in the structure of L-alanyl-glycyl-glycine (32).

Ala–Gly–Gly. All these polymers have been proposed to have a triple helical structure in which each individual chain is a γ -helix (26–30).

In oligopeptides, the γ -helix is observed whenever the sequence has a predominance of prolines in the sequence, but the molecular packing is not reminiscent of triple-helical structures. Our analysis of the crystal structure of L-alanyl-glycyl-glycine (31, 32), therefore, presents some interesting features. The molecule adopts a γ -helix conformation with $\phi, \psi = -83^\circ, 170^\circ$. The crystal packing is also reminiscent of triple-helical structures (collagen-fold) (figure 8).

Till recently, the occurrence of the collagen-fold was believed to be due to the presence of a large number of prolines in the sequence $(\text{Gly-x-y})_n$, characteristic of collagen polymers. Recent FT–IR and Raman spectroscopic studies on the tripeptide (L-alanyl-glycyl-glycine) show similarities with collagen (33). Our studies on the crystal structure of L-alanyl-glycyl-glycine thus provides an opportunity to examine the nature of molecular interactions which give rise to the collagen fold. As in collagen, each peptide molecule in the crystal is surrounded by six neighbours in a pseudo-hexagonal packing arrangement, and linked by two interchain hydrogen bonds nearly perpendicular to the chain direction. These interchain hydrogen bonds have lengths of 2.93 and 2.99 Å, somewhat longer than the value of 2.8 Å proposed for polyglycine II from model-building studies (29). The packing arrangement also introduces two short contacts $2\text{C}^\alpha - 1\text{O}' = 3.30$ Å and $3\text{C}^\alpha - 1\text{O}' = 3.25$ Å and the corresponding H atoms make $\text{H}-\text{C}^\alpha-\text{O}'$ angles of 66° and 51° respectively, indicating the presence of $\text{CH}\cdots\text{O}$ interactions. In fact, these features were also indicated by infra-red studies on polyglycine II (34), and were incorporated in the models proposed by Ramachandran *et al.* (35).

Folded Conformations

Backbone conformations which do not belong to either helical or sheet structure are referred to as folded conformations. The most important are the reverse turn conformations which provide stable hydrogen-bonded arrangements and give rise to a reversal of the chain direction in the overall folding pattern of a peptide or protein chain.

A reverse turn occurs when the peptide backbone twists at two successive C^α atoms in such a way that the peptide carbonyl group of a residue comes within hydrogen

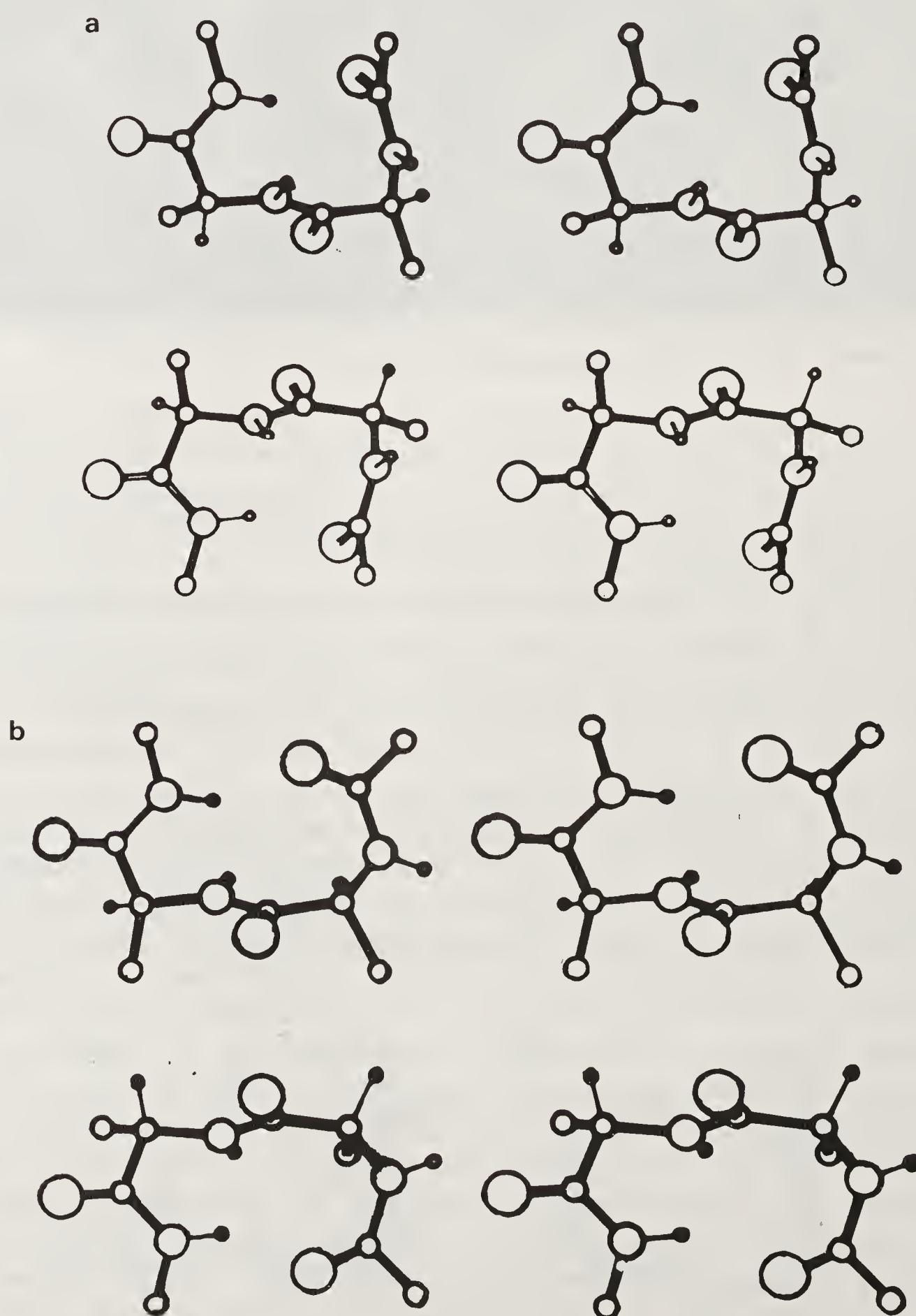


Figure 9. (Continued)

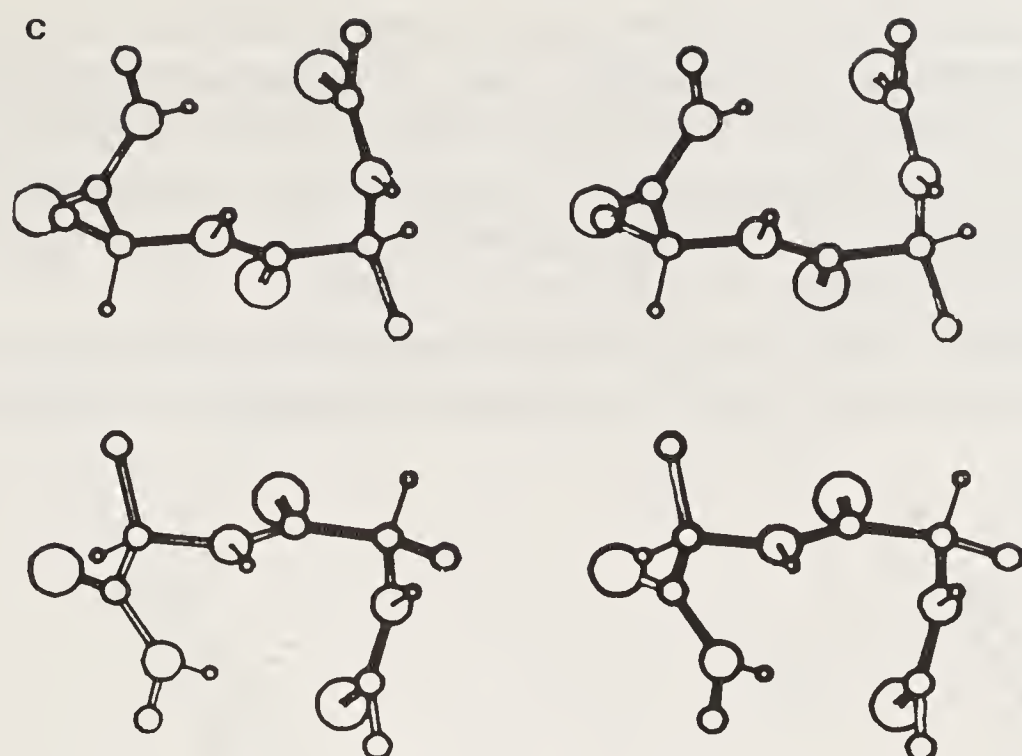


Figure 9. Stereoscopic view of reverse turn conformations: **a.** Type I (top) and Type I' (bottom); **b.** Type II (top) and Type II' (bottom); **c.** Type III (top) and Type III' (bottom).

bonding distance of the amide group of a neighbouring residue. The Madras School, led by Professor Ramachandran predicted the existence of the reverse turns long before they were observed in peptides and proteins (36).

Of the various types of reverse turn conformations resulting in such intramolecular hydrogen bonds, the ones with $4 \rightarrow 1$ hydrogen bonds (i.e., between the NH group of the $i + 3$ residue and the $C=O$ group of the i^{th} residue) were predicted to be the most stable (36). In these stable arrangements, there are 3 consecutive peptide units enclosed between four C^α atoms (numbered C_1^α , C_2^α , C_3^α and C_4^α) and a $4 \rightarrow 1$ hydrogen bond occurs whenever the (ϕ, ψ) values at C_2^α and C_3^α have the values given in table 2.

The mirror-image conformations of the above, corresponding to a sign reversal for all the torsion angles, also produce the same hydrogen-bonded interactions, and are referred to as Type I', II' and III' turns. Figure 9 shows a stereoscopic view of the various reverse turn conformations.

The difference between Types I and II consists in the middle peptide unit (between C_2^α and C_3^α) flipping over by nearly 180° . Also, in type II, steric hindrance between carbonyl oxygen (O_2') and the side chain on C_3^α requires that residue 3 be glycine

Table 2. Values of ϕ, ψ at C_2^α and C_3^α .

	ϕ_2, ψ_2	ϕ_3, ψ_3
Type I	$-60, -30$	$-90, 0$
II	$-60, 120$	$80, 0$
III	$-60, -30$	$-60, -30$

or a D-amino acid. For this reason, type I turns are known as LL turns while type II turns are referred to as LD turns, the L and D referring to the chirality of the residue at C_2^α and C_3^α . Clearly, type I' turns will require glycines at both C_2^α and C_3^α while Type II' turns will have a glycine at C_2^α . Type III turn is a fragment of a right-handed 3_{10} helix. Comparison of the bend angles shows that type I turn can, in fact, be considered to be a fragment of a distorted 3_{10} helix.

In oligopeptides, Type I turn occurs more frequently. In these cases, steric repulsion appears to make the angle $N_2C_2^\alpha C_2'$ significantly larger than the mean value of 110° .

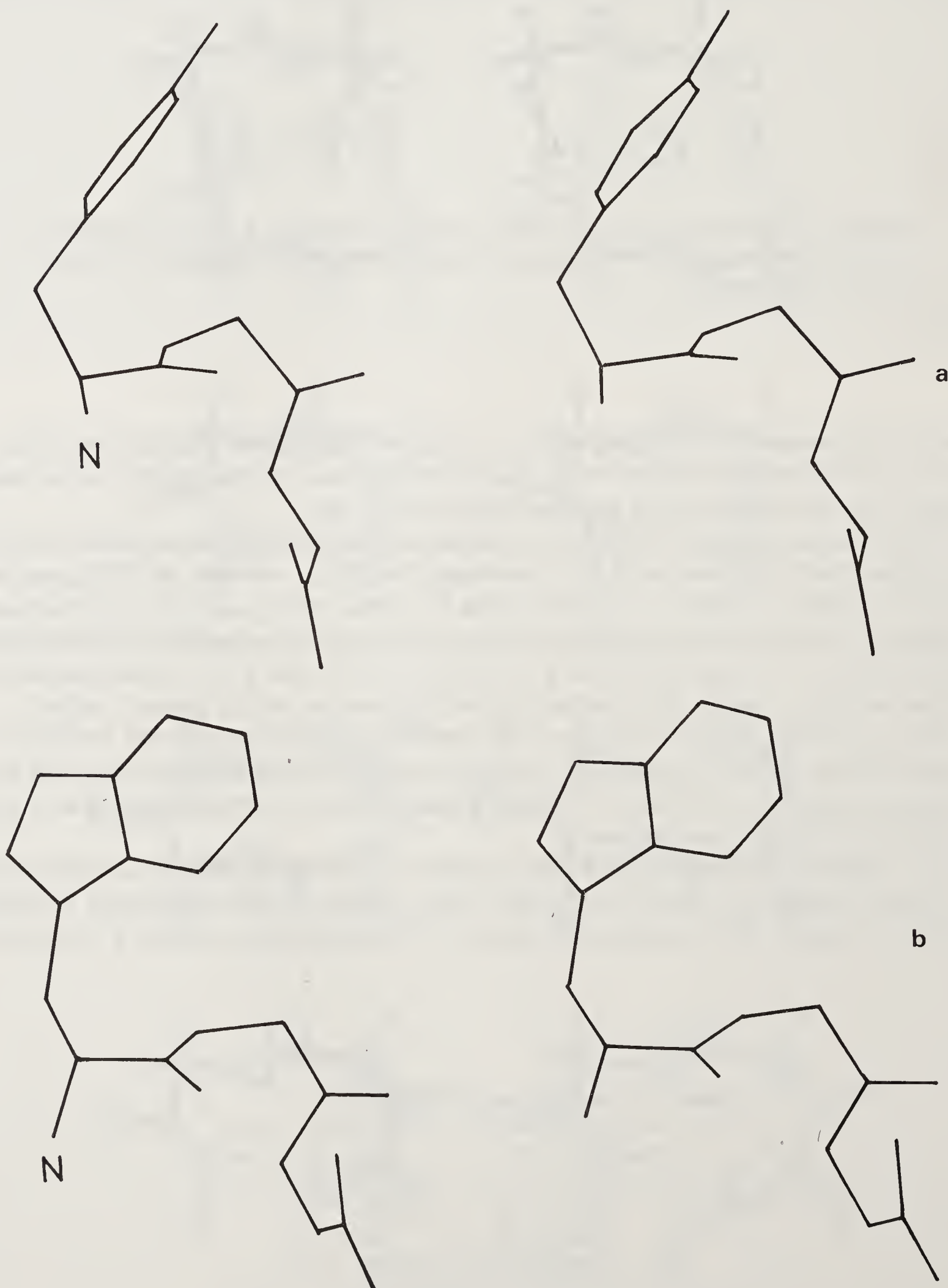


Figure 10. (Continued)

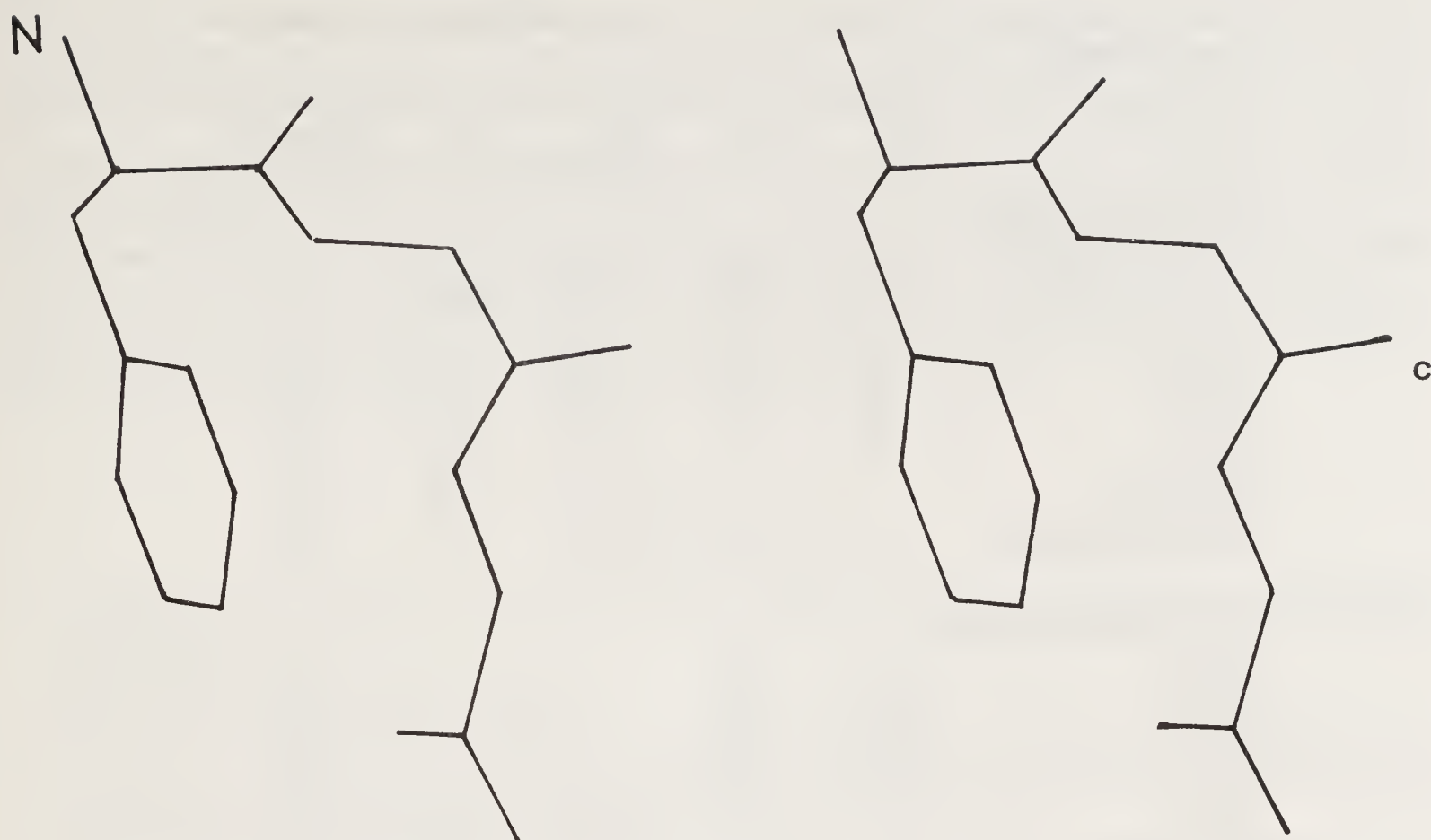


Figure 10. Comparison of the conformations of three tripeptides of the type X-Gly-Gly where X is an aromatic residue. **a.** Tyrosyl-glycyl-glycine (39); **b.** Tryptophanyl-glycyl-glycine (38); **c.** Phenylalanyl-glycyl-glycine (37).

Some General Sequence-Dependent Features

Comparison of the X-ray structures of two tripeptides, phenylalanyl-glycyl-glycine (37) and tryptophanyl-glycyl-glycine (38) with that of tyrosyl-glycyl-glycine (39) reveals the following features (figure 10). All are of the type X-Gly-Gly where X is an aromatic residue, all crystallize from aqueous solution at room temperature in the same spacegroup $P2_12_12_1$, with one molecule per asymmetric unit. When $X = \text{Tyr}$ or Trp , the peptide backbone folding is essentially similar, with $\phi, \psi = (85^\circ, 10^\circ)$; but when $X = \text{Phe}$, the folding is different, with $\phi, \psi = (-90^\circ, -30^\circ)$. Similarly, when $X = \text{Tyr}$ or Trp , the packing of the molecules in the lattice includes one or two water molecules of crystallization, but not when $X = \text{Phe}$. The implications are that in terms of side-chain hydrophobicity, tryptophan is quite similar to tyrosine, but phenylalanine is not. That is, while phenylalanine is quite hydrophobic, tryptophan and tyrosine are more polar, and hence the folding of peptide chains in the neighbourhood of aromatic residues may also be influenced by this differential polarity. The substitution of Phe for Tyr or *vice versa* in a protein chain may therefore give rise to significantly different local conformations, and hence need not be 'conservative'.

The author's studies on several tripeptides possessing two contiguous glycines in the sequence lead to the observation that a non-terminal glycine frequently adopts a "D-residue" (ϕ is positive) conformation whenever the preceding or following residue is aromatic. When the residue involved is non-aromatic, this tendency of glycine is the least. Table 3 compares the observed backbone torsion angles ϕ and ψ for the two contiguous glycines in several peptides to highlight this point.

Table 3. Backbone torsion angles for two contiguous glycines in several peptides

Peptide	phi	psi	phi	psi	Reference
<i>Aromatic peptides:</i>					
YGGFL: Mol. B	151	−155	154	−151	7
Mol. C	141	−157	174	−170	7
YGGFL	59	25	97	−7	41
YGGF	74	11	91	0	42
YGG	81	12	−104	−153	39
WGG	88	9	−85	−178	38
GGF·HCl			64	−154	43
t-boc GGF OEt	98	9	109	62	44
<i>Non-aromatic peptides:</i>					
AGG	−83	169			45
VGG	−155	155			14
IGG	−160	163			15
PGG	−71	167			46
GGI			−106	−151	47
GGV			−77	−22	24

The significance and structural basis of this behaviour on the part of glycine needs to be explored further.

Epilogue

It is a pleasure to contribute this article as a tribute to Professor G.N. Ramachandran who was instrumental in developing the school of conformational analysis of biopolymers at Madras nearly thirty years ago. Under his guidance, the Madras School was also instrumental in carrying out X-ray crystallographic studies on many amino acids, along with the seminal work on the Ramachandran Diagram which involved modelling studies using planar peptide units. At that time, the number of peptide molecules whose crystal structures were known, even at the international level, could be counted by the fingers of one's hands. Professor Ramachandran wished that the Madras School would get involved with X-ray studies on peptides. As luck would have it, he got a small sample of a dipeptide from the late Professor L. K. Ramachandran who was then at the Biochemistry Department of the University of Madras (it was literally next door). Professor GNR gave the dipeptide sample to me, indicating that I should try to crystallize it. There was hardly a few milligrams of the dipeptide (it was N-methyl-DL-leucyl-glycine), enough for only one crystallization-trial! In those days, direct methods of structure analysis were unheard of, and no one did crystallography on a molecule that did not have a "heavy atom" in the chemical formula! I emptied the entire peptide sample onto a watch-glass, added concentrated hydrobromic acid, and set the thing aside to evaporate. After a week, I found just one crystal! The state of ecstasy that a researcher goes through when he gets crystals of an important and interesting compound, that too at the very first attempt, cannot be described in a few words. My sense of jubilation was only tempered by the fact

that I had only one crystal. Sitting on a work-bench, I got the crystal mounted on a goniometer-head, and as I was picking up the goniometer-head, suddenly a loud noise distracted me and before I realized it, the crystal had dropped to the floor. I went down on my knees and searched the floor very carefully, looking for my lone precious peptide-crystal amidst other junk. Finally I did find the crystal, and went ahead and characterized it! The structure determination of that peptide (40) represents the first X-ray study of a peptide in India nearly 25 years ago, and it is a tribute to Professor Ramachandran's perspicacity that peptide structure analysis continues to be an active area of research activity at Madras even today.

References

1. Anfinsen, C. B. (1964) in *New Perspectives in Biology*, ed. Sela, M. (Elsevier, Amsterdam), pp. 42–50.
2. Brookhaven Protein Data Bank.
3. Pauling, L., Corey, R. B. & Branson, H. R. (1951) *Proc. Natl. Acad. Sci. U.S.A.* **37**, 205–211.
4. Karle, I. L. (1981) in *The Peptides: Analysis, Synthesis and Biology*, eds. Gross, E. & Meienhofer, J. (Academic Press, New York), Vol. 4, pp. 1–55.
5. Hruby, Victor J. (1985) in *The Peptides*, (Academic Press, New York), Vol. 7, pp. 1–11.
6. IUPAC-IUB Commission on Biochemical Nomenclature (1970) *J. Mol. Biol.* **52**, 1–17.
7. Karle, I. L., Karle, J., Mastropaolo, D., Camerman, A. & Camerman, N. (1983) *Acta Crystallogr.* **B39**, 625–637.
8. Griffin, J. F., Langs, D. A., Smith, G. D., Blundell, T. L., Tickle, I. J. & Bedarkar, S. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3272–3276.
9. Stezowski, J. J., Eckle, E. & Bajusz, S. (1985) *J. Chem. Soc., Chem. Commun.* Number 11, 681–682.
10. Doi, M., Ishida, T., Inoue, M., Fujiwara, T., Tomita, K., Kimura, T. & Sakakibara, S. (1984) *FEBS Lett.* **170** (2), 229–231.
11. Doi, M., Tanaka, M., Ishida, T., Inoue, M., Fujiwara, T., Tomita, K., Kimura, T., Sakakibara, S. Sheldrick, G. M. (1987) *J. Biochem (Tokyo)* **101**, 485–490.
12. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein. Chem.* **23**, 283–438.
13. Marsh, R. E. & Glusker, J. P. (1961) *Acta Crystallogr.* **14**, 1110–1116.
14. Lalitha, V., Murali, R. & Subramanian, E. (1986) *Int. J. Pept. Protein Res.* **27**, 472–477.
15. Subramanian, E., unpublished.
16. Bandekar, S. & Krimm, S. (1988) *Biopolymers* **27**, 885–908.
17. Arnott, S., Dover, S. D. & Elliott, E. (1967) *J. Mol. Biol.* **30**, 209–212.
18. Subramanian, E. & Parthasarathy, R. (1987) *Curr. Sci.* **56**, 1210–1213.
19. Lalitha, V. & Subramanian, E. (1982) *Cryst. Struct. Commun.* **11**, 561–564.
20. Srikrishnan, T., Winiewicz, N. & Parthasarathy, R. (1982) *Int. J. Pept. Protein Res.* **19**, 103–113.
21. Fawcett, J. K., Camerman, N. & Camerman, A. (1975) *Acta Crystallogr.* **B31**, 658–666.
22. Subramanian, E. & Parthasarathy, R. (1989) *Int. J. Pept. Protein Res.* **33**, 345–347.
23. Karle, I. L., Flippen-Anderson, J. L., Sukumar, M. & Balaram, P. (1990) *Int. J. Pept. Protein Res.* **35**, 518–526.
24. Lalitha, V., Subramanian, E. & Bordner, J. (1984) *Int. J. Pept. Protein Res.* **24**, 437–441.
25. Ramachandran, G. N. (1981) in *Biomolecular Stereodynamics*, ed. Sarma, R. H. (Adenine Press) Vol. 2, pp. 1–36.
26. Ramachandran, G. N. & Kartha, G. (1954) *Nature (London)* **174**, 269–270.
27. Ramachandran, G. N. & Kartha, G. (1955) *Nature (London)* **176**, 593–595.
28. Cowan, P. M. & McGavin, S. (1955) *Nature (London)* **176**, 501–503.
29. Crick, F. H. C. & Rich, A. (1955) *Nature (London)* **176**, 780–781.
30. Andrias, J. C., Anderson, J. M. & Walton, A. G. (1971) *Biopolymers* **10**, 1049–1057.
31. Subramanian, E. (1981) *Curr. Sci.* **50**, 347–348.
32. Subramanian, E. & Lalitha, V. (1983) *Biopolymers* **22**, 833–838.

33. Renugopalakrishnan, V., Kloumann, P. H. B. & Bhatnagar, R. S. (1984) *Biopolymers* **23**, 623–627.
34. Krimm, S., Kuroiwa, K. & Rebane, T. (1967) in *Conformation of Biopolymers*, ed. Ramachandran, G. N. (Academic Press, New York), Vol. 2, pp. 439–447.
35. Ramachandran, G. N., Ramakrishnan, C. & Venkatachalam, C. M. (1967) in *Conformation of Biopolymers*, ed. Ramachandran, G. N. (Academic Press, New York), Vol. 2, pp. 429–438.
36. Venkatachalam, C. M. (1968) *Biopolymers* **6**, 1425–1436.
37. Subramanian, E. & Sahayamary, J. J. (1989) *Int. J. Pept. Protein Res.* **34**, 211–214.
38. Subramanian, E. & Sahayamary, J. J. (1989) *Int. J. Pept. Protein Res.* **34**, 134–138.
39. Carson, W. M. & Hackert, M. L. (1978) *Acta Crystallogr.* **B34**, 1275–1280.
40. Chandrasekharan, R. & Subramanian, E. (1969). *Acta Crystallogr.* **B25**, 2599–2604.
41. Smith, D. & Griffin, J. (1978) *Science* **199**, 1214–1216.
42. Prange, P. T. & Pascard, C. (1979) *Acta Crystallogr.* **B35** 1812–1819.
43. Murali, R & Subramanian, E. (1987) *Int. J. Pept. Protein Res.* **29**, 374–380.
44. Ishida, T., Tanabe, N. & Inoue, M. (1983) *Acta Crystallogr.* **C39**, 110–112.
45. Lalitha, V. & Subramanian, E. (1985) *Indian J. Pure Appl. Phys.* **23**, 506–508.
46. Lalitha, V., Subramanian, E. & Parthasarathy, R. (1986) *Int. J. Pept. Protein Res.* **27**, 223–228.
47. Lalitha, V., Subramanian, E. & Bordner, J. (1984) *Int. J. Pept. Protein Res.* **24**, 123–128.

Thermally Driven Self-Assembly, Molecular Structuring and Entropic Mechanisms in Elastomeric Polypeptides

Dan W. Urry

Laboratory of Molecular Biophysics, The University of Alabama at Birmingham,
P.O. Box 300, University Station, Birmingham, Alabama 35294, USA

Introduction

The presence of repeating sequences in the primary structure of the mammalian elastic protein, elastin (1–5), has provided a unique opportunity to develop a detailed understanding of biological entropic elasticity (6, 7); it has afforded special insight into the driving forces of protein folding, assembly and function (8, 9), particularly as related to chemomechanical transduction (10, 11); and it has resulted in a new class of biomaterials (12, 13), bioelastic materials, which constitute an early emerging subset of what are now being referred to as protein-based polymers. Major repeating sequences presently known can be written: (Val 1–Pro 2–Gly 3–Val 4–Gly 5)_n or poly(VPGVG), (Ala 1–Pro 2–Gly 3–Val 4–Gly 5–Val 6)_n or poly(APGVGV), (Val 1–Pro 2–Gly 3–Phe 4–Gly 5–Val 6–Gly 7–Ala 8–Gly 9)_n or poly(VPGFGVGAG), (Val 1–Pro 2–Gly 3–Gly 4) or poly(VPGG), (Val 1–Pro 2–Gly 3)_n or poly(VPG), and (Gly 1–Val 2–Gly 3–Gly 4–Leu 5)_n or poly(GVGGL). These and many analogs have been prepared chemically. Of these, poly(VPGVG), poly(VPGFGVGAG), and poly(VPGG) on γ -irradiation cross-linking have been shown to be elastic whereas the other repeats have not been demonstrated to be elastic.

Many physical methods have been used in studying the repeating sequences and their analogs, for example, electron microscopy and optical diffraction, circular dichroism, nuclear magnetic resonance for both conformational and relaxational characterization, X-ray diffraction, Raman scattering, dielectric relaxation, and mechanical characterizations (6, 14–16). The studies of physical properties demonstrate the presence of non-random structure and particularly an increase in intramolecular and intermolecular order on raising the temperature to the functional physiological range. In spite of this demonstrated absence of random networks, traditional analyses argue for a dominance of entropic elastomeric force in the cross-linked matrices.

In what follows will be reviewed studies which demonstrate that the amphiphilic balance of these sequential polypeptides is such that they self-assemble into more-ordered molecular assemblies on raising the temperature, that is, they exhibit inverse temperature transitions. Then the nature of the elasticity in self-assembling systems will be considered in terms of molecular processes occurring within different

temperature ranges which relate differentially to the temperature dependence of elastomeric force. Finally, a brief review of the nature of entropic elastomeric force in these self-assembling (non-random) systems will be given.

Self-assembly by Means of an Inverse Temperature Transition

Thermally Driven Self-Assembly

Self-assembly of the repeating hexapeptide: When polypeptides, dissolved in water at low temperatures, have the requisite amphiphilicity, that is, the appropriate balance of apolar (hydrophobic) and polar moieties, on raising the temperature they associate



Figure 1. Thermally driven reversible crystallization of cyclo(APGVGV)₂. This cyclic analog of the polyhexapeptide of elastin, poly(APGVGV), is soluble in water at low temperature, crystallizes on raising the temperature, and redissolves on lowering the temperature. (Reproduced with permission from reference 17).

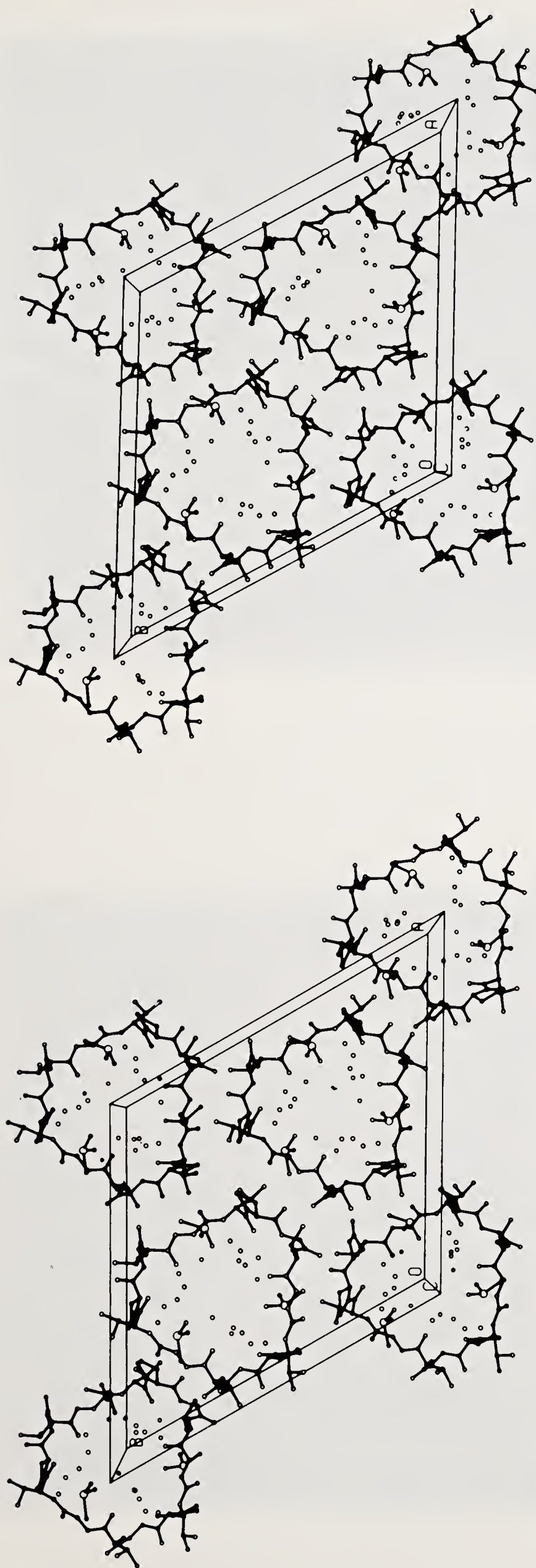
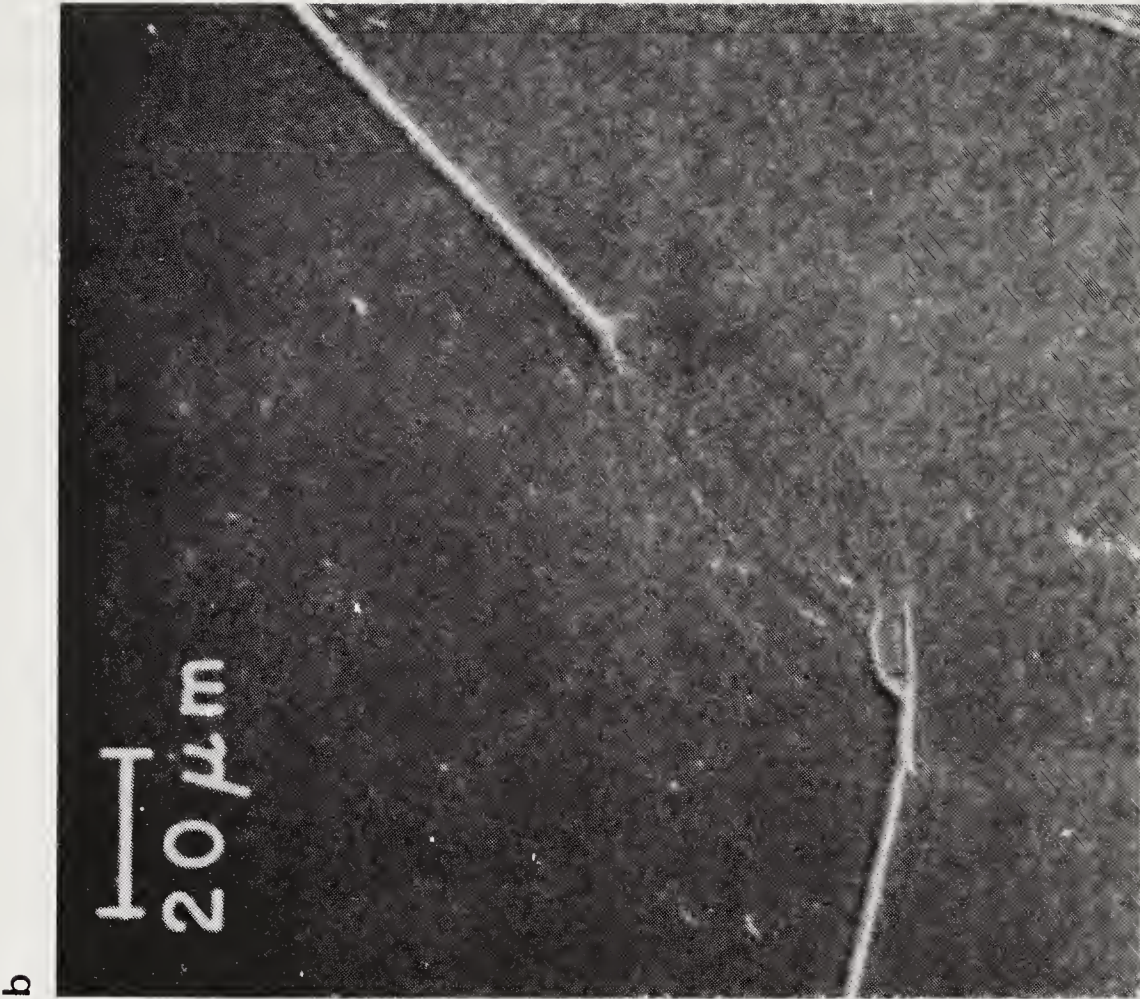
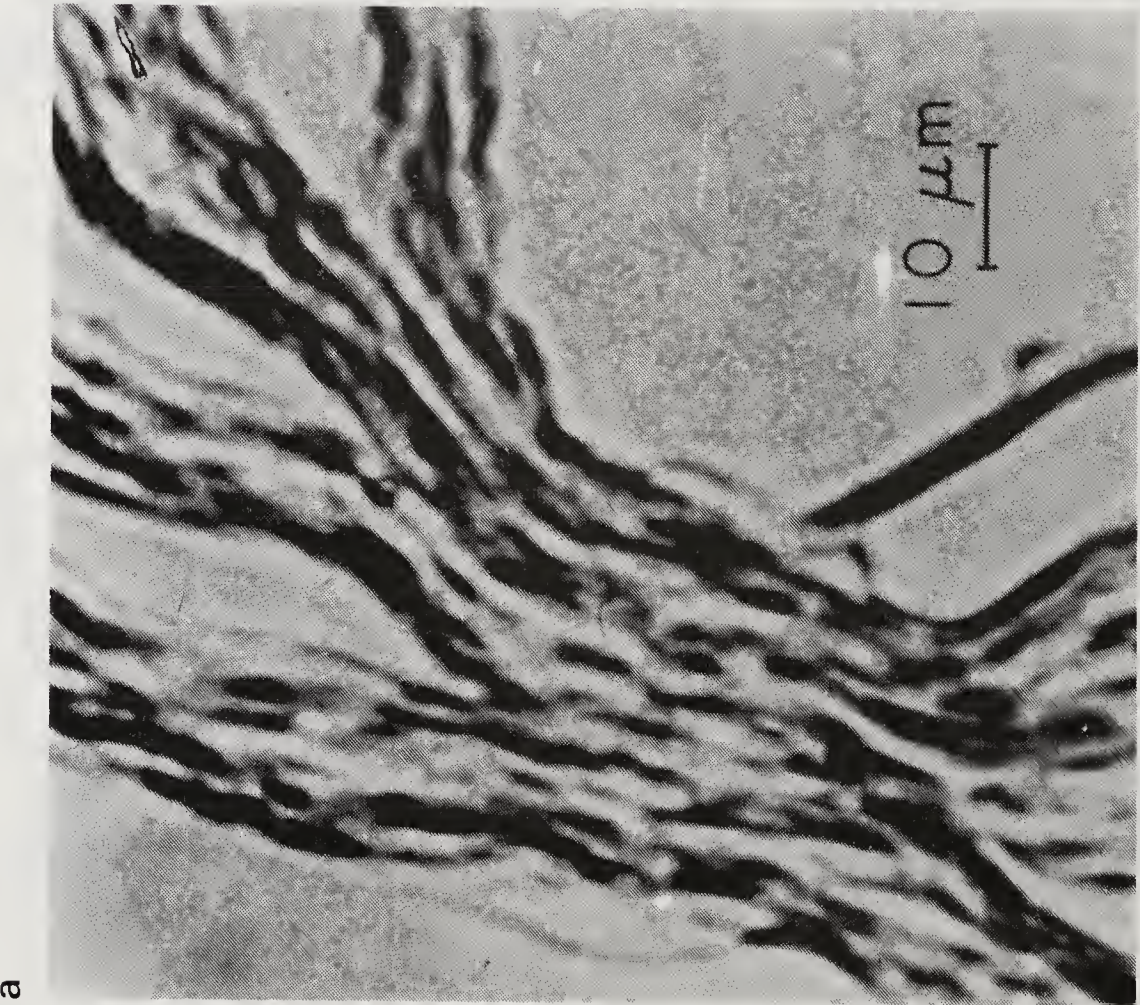


Figure 2. Crystal structure of cyclo(VPGVG)₃ the cyclic conformational correlate of the polypeptide of elastin, poly(VPGVG). The dots within the cyclic structures are approximate positions of water molecules. An enlarged stereo pair of the molecular structure is given in figure 9a. (Reproduced with permission from reference 19).



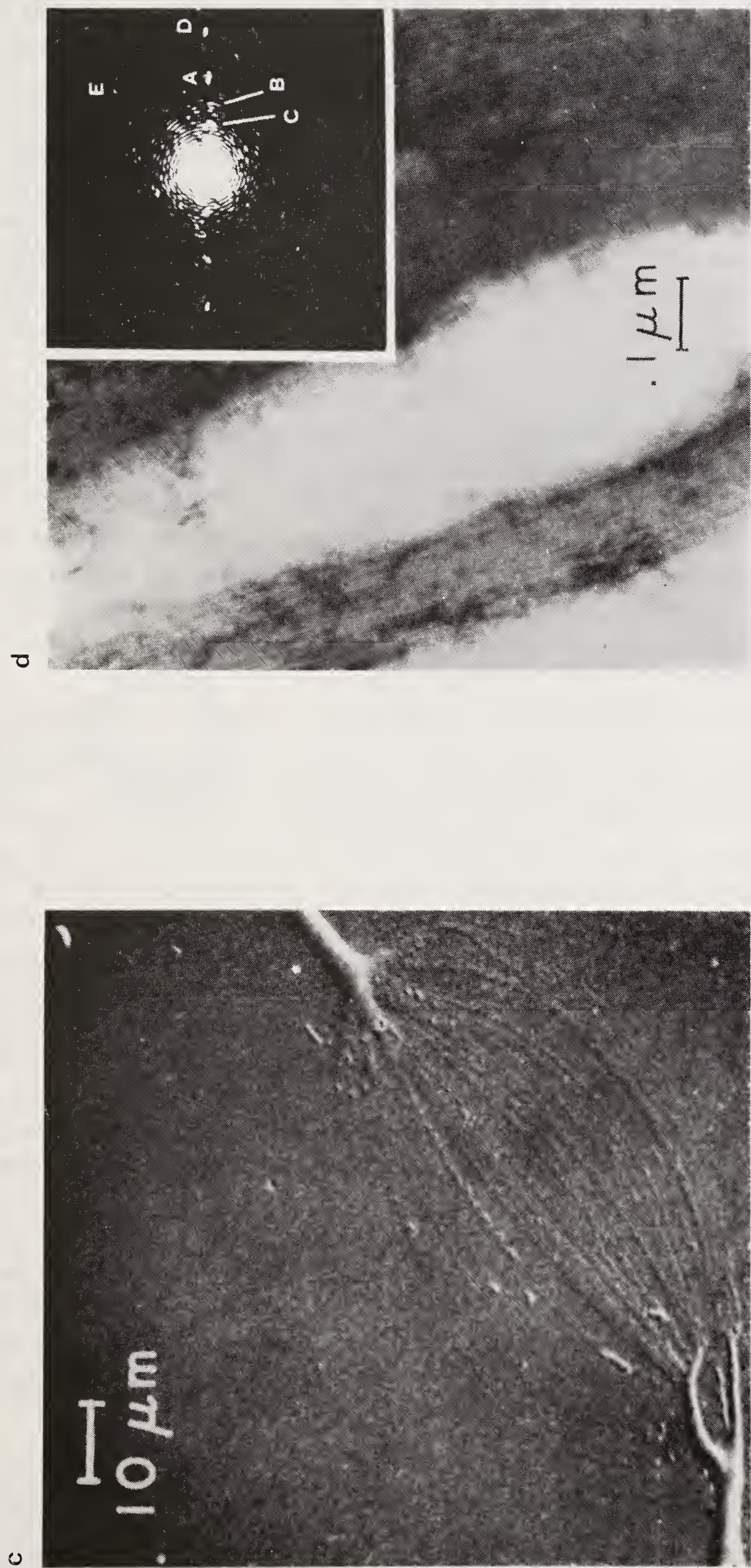


Figure 3. Self-assembly of the polypeptide of elastin, a thermally driven association to form anisotropic fibers. See text for discussion. (a reproduced with permission from reference 21; b and c reproduced with permission from reference 20, and d adapted with permission from references 22 and 23).

to form more-ordered structural states. This has been most unambiguously demonstrated by a cyclic analog of poly(APGVGV). In particular, cyclo(APGVGV)₂ is soluble in water below physiological temperatures but on raising the temperature the molecules go from being dispersed and randomly related in solution to associating in the formation of crystals (17). Crystals of cyclo(APGVGV)₂ obtained simply by raising the temperature are shown in figure 1 and determination of the crystal structure is in progress (18). It is unequivocal that cyclo(APGVGV)₂ exhibits an inverse temperature transition resulting in increased order on raising the temperature. In a similar manner with a similar concentration dependence for the temperature profile for aggregation, poly(APGVGV) aggregates on raising the temperature. This, of course, is to be expected since high molecular weight poly(APGVGV) and cyclo(APGVGV)₂ have essentially identical amphiphilicity.

Self-assembly of the repeating pentapeptide: The cyclopentadecapeptide analog of poly(VPGVG), i.e., cyclo(VPGVG)₃, also is soluble at low temperature and also aggregates on raising the temperature. The crystal structure of cyclo(VPGVG)₃ has been determined as shown in figure 2 (19). Interestingly, all of the substantial water in the crystal is within the stack of cyclic molecules; no water is found between molecules; and the intermolecular interactions are dominated by hydrophobic contacts between molecules within each stack and between stacks of molecules. Again, the linear high polymer poly(VPGVG) is soluble in water at low temperature, and it also aggregates on raising the temperature (14).

Thermally driven folding and assembly into fibers: The aggregation of poly(VPGVG) can be followed in an informative way when occasional lysine (K) residues are introduced in one synthesis, occasional glutamic acid (E) residues are introduced in a second synthesis, and the aggregation is thermally initiated in the presence of a water soluble carbodiimide which will cross-link the Glu and Lys side chains whenever a polypentapeptide is added to the aggregate with the Glu and Lys side chains in juxtaposition. When this is done the polypentapeptide is seen to self-assemble into fibers (see figure 3) which can be observed in the light microscope without any fixative utilizing only form birefringence (20–22); in the scanning electron microscope with an aluminum coating the fibers can be seen to splay-out into fibrils and to recoalesce back into the same sized fiber, and in the transmission electron microscope negatively stained fibrils can be seen to be comprised of parallel aligned filaments with a twisted substructure (14). Furthermore, optical diffraction of the TEM micrograph shows both equatorial and off-equatorial reflections relevant to the dimensions of the twisted filament (23).

These self-assembled fibers have been found to be elastomeric (20), and complete phase diagrams showing the temperature and concentration dependence of the cloud point (coacervation) and spinodal lines have been developed using elastic and quasi-elastic light scattering which support a description of the very early stages of the self-assembly of the bioelastomeric structures as resulting from density fluctuations (24–26).

Molecular Structure of the Thermally Assembled State of Poly(VPGVG)

In the very earliest conformational studies of poly(VPGVG) the β -turn was shown to be the dominant secondary structural feature (27–29) and specifically that it was the Type II Pro 2–Gly 3 β -turn (30) as shown in figure 4a. It was then anticipated

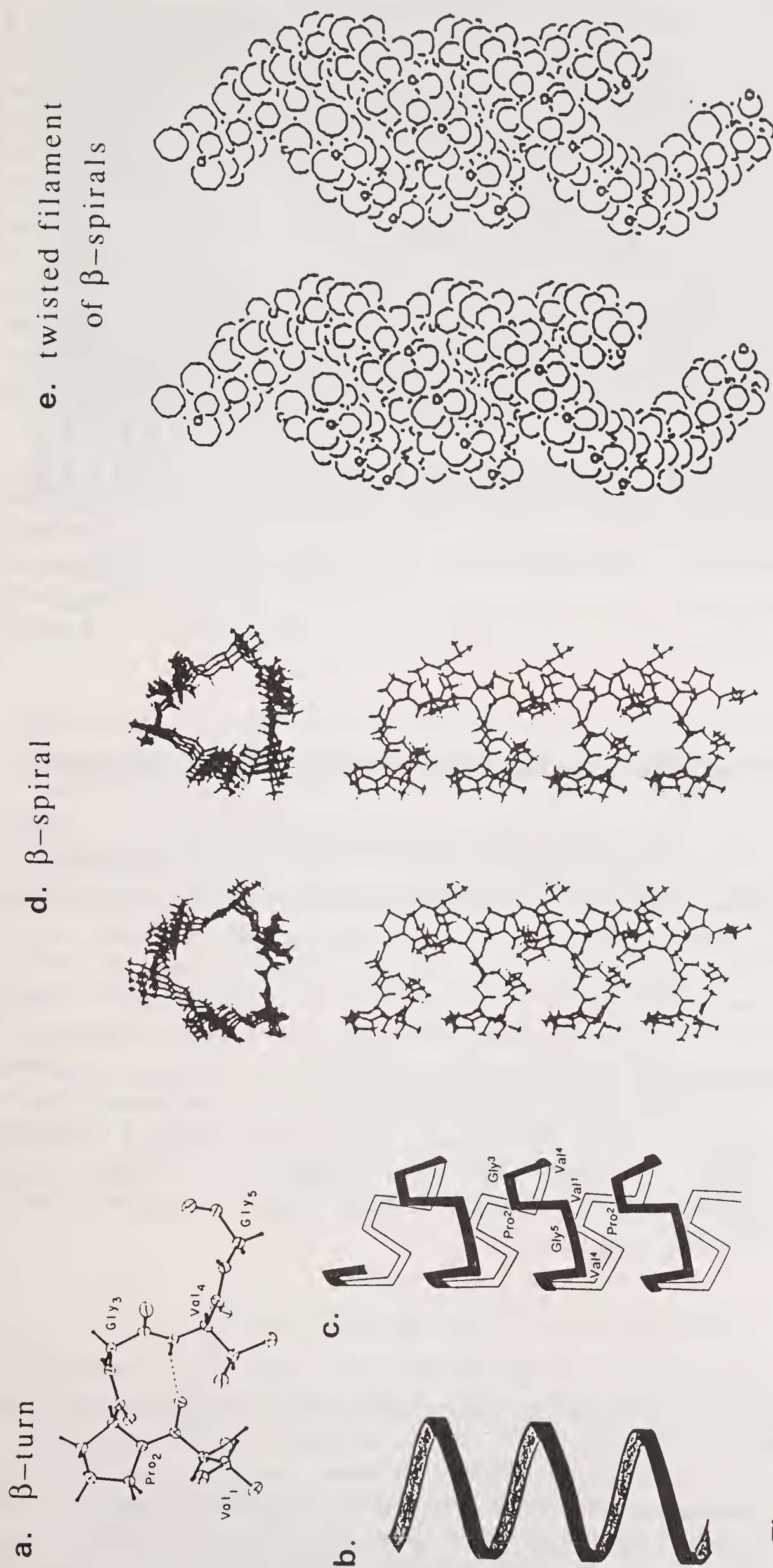


Figure 4. Molecular structure of poly(VPGVG) (**a**) The β -turn with Pro 2 and Gly 3 at the corners, with a hydrogen bond between the Val 1 C-O and the Val 4 NH, and with the end Pro 2-Gly 3 peptide moiety with the C-O directed upward out of the plane. (**b**) Schematic helix (c) Schematic helix with a schematic representation of the β -turns functioning as spacers between the turns of the helix. (**d**) Stereo pair of the detailed β -spiral, a helical structure with 2.9 pentamers per turn; upper, axis view; lower side view showing the interturn hydrophobic contacts and the Val 4-Gly 5-Val 1 suspended segments. (**e**) Stereo pair of three β -spirals supercoiling to form a twisted filament. (**a** reproduced with permission from reference 19; **b** reproduced with permission from reference 21; **c** and **d** reproduced with permission from reference 12; and **e** reproduced with permission from reference 37).

and found that cyclo(VPGVG)₃ had essentially the same conformation as the linear poly(VPGVG), that is, that there were approximately three pentamers per turn in a helical structure (15, 31) called a β -spiral. The crystal structure of cyclo(VPGVG)₃ confirmed the presence of the Type II β -turn (19) and with the solution conformation of cyclo(VPGVG)₃ (32) provided the basis for a detailed computed structure of the β -spiral of poly(VPGVG) (33) which had 2.7 pentamers per turn. Two dimensional nuclear magnetic resonance studies comparing cyclo(VPGVG)₃ and poly(VPGVG) suggested that there were closer to three pentamers per turn (34) due to a suggested $\text{Val}_i^1\gamma\text{CH}_3 \leftrightarrow \text{Pro}_{i+x}^2\beta\text{CH}_2$ proton-proton interaction which was confirmed by synthesis and nuclear Overhauser effect studies on $\text{Val}^1(\text{d}_8)$ -poly(VPGVG) and $\text{Val}^4(\text{d}_8)$ -poly(VPGVG) and by computations (35). Details of the interturn interactions are apparent in figures 4c and 4d (12). The result of three pentamers/turn was independently obtained by Wasserman and Salemme from molecular dynamics calculations in water (36).

The molecular description therefore of poly(VPGVG) is a series of Pro 2–Gly 3 Type II β -turns which on raising the temperature wrap-up into a helix optimizing the interturn hydrophobic contacts with the β -turns functioning as spacers between turns of the helix. Several of these β -spirals then supercoil to form twisted filaments (37) of the dimensions seen in the transmission electron micrographs of negatively stained aggregates that were yet small enough for Brownian motion to keep in suspension.

Nature of Elasticity in Self-Assembling Protein-Based Polymers

The above brief sketch shows these elastic-protein-based polymers to be of such an amphiphilic nature as to self-assemble into more-ordered states on increasing the temperature of aqueous solutions. Now it will be seen how this thermally driven molecular structure formation relates to the thermoelastic properties of macroscopic matrices prepared from poly(VPGVG). On raising the temperature from 25° to 37°C, the aggregation of poly(VPGVG) in water results in a phase transition, called coacervation, in which the peptide-rich phase, the coacervate, is approximately 40% peptide, 60% water by weight. This viscoelastic, glue-like coacervate can be formed into band shapes and 20 Mrad γ -irradiation cross-linked to form elastomeric bands referred to as X²⁰-poly(VPGVG). A representative band is seen in figure 5. The effect of 20 Mrad γ -irradiation on primary structure is so minimal that extraneous peaks are difficult to detect in carbon-13 and nitrogen-15 nuclear magnetic resonance spectra of enriched samples (38, 39).

Two Stages of the Thermoelasticity Data

The behaviour of a band of X²⁰-poly(VPGVG) in response to increases in temperature can be conveniently studied under either of two conditions: temperature dependence of length at constant load, $(\partial l/\partial T)_f$, or temperature dependence of force at constant length, $(\partial f/\partial T)_l$. Characterization of the former (see figure 6) demonstrates thermomechanical transduction and the efficiency of the process (40) whereas analysis of the latter provides insight into the nature of the elastomeric force wherein the data is usefully plotted as $\ln(f/T)$ vs. T as shown in figure 7 (41). In both cases the data

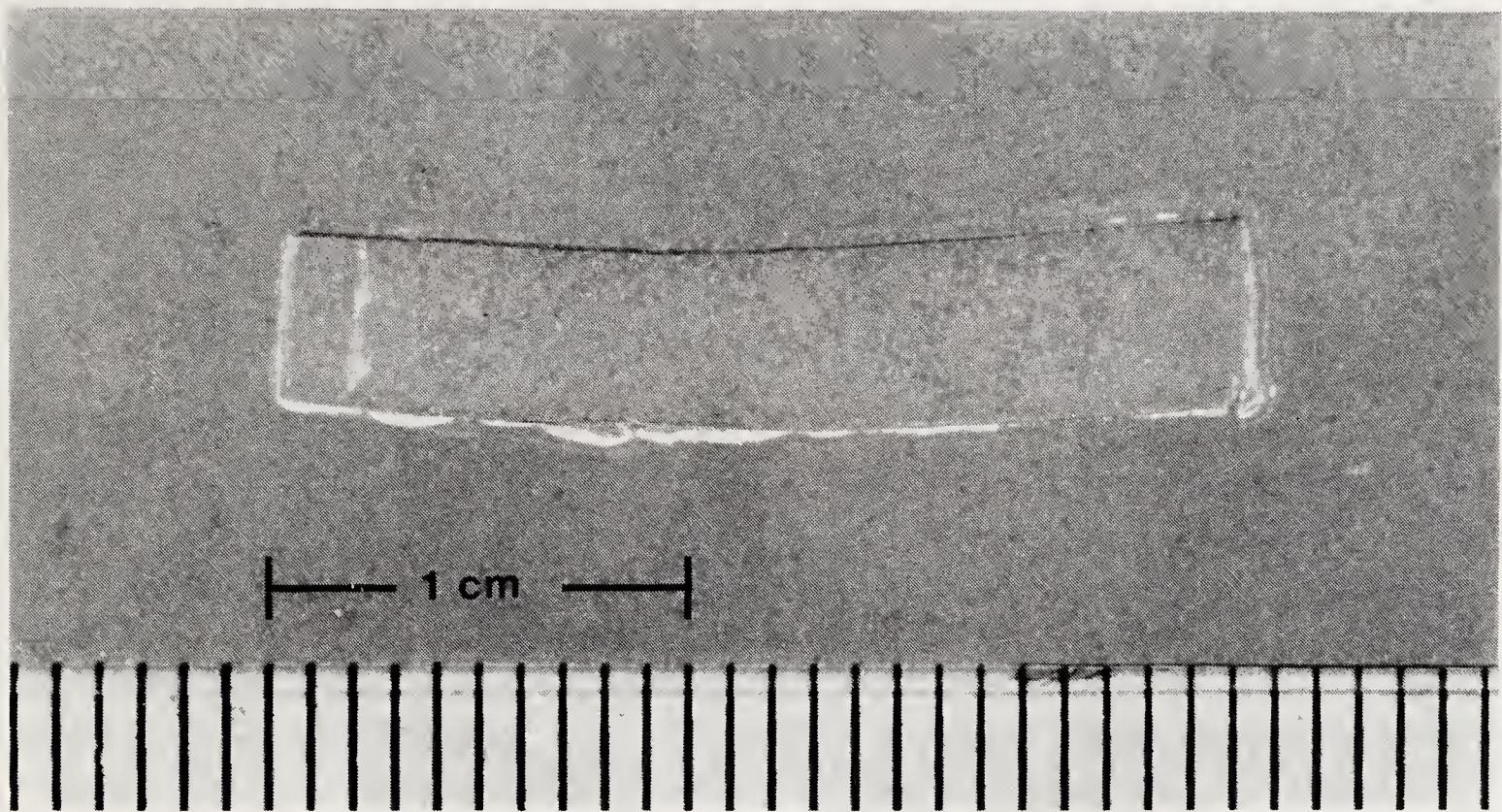


Figure 5. An elastomeric strip of 20 Mrad γ -irradiation cross-linked poly(VPGVG) designated as X^{20} -poly(VPGVG).

demonstrate a transition region with a dramatic decrease in length or increase in force as the temperature is raised through the range of the inverse temperature transition and in both cases once the transition is complete there is limited further change with increasing temperature. In what follows the two temperature processes will be considered. The lower temperature process which is the inverse temperature transition of folding and assembly will be considered in terms of a review of the

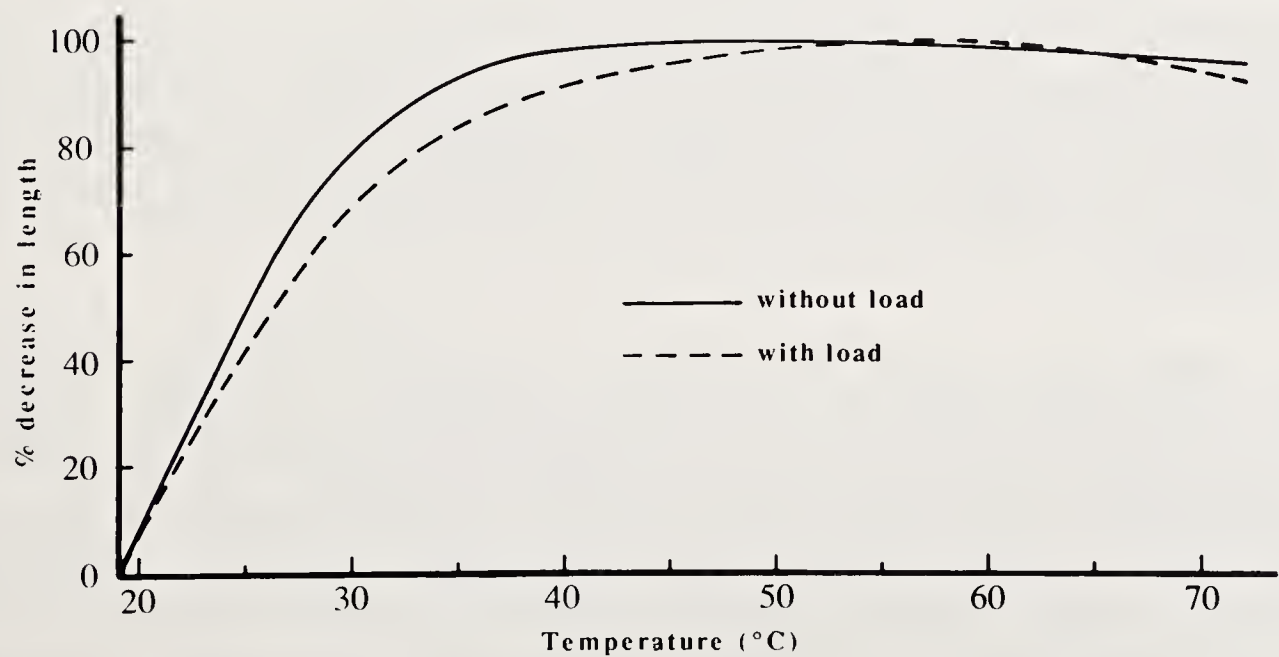


Figure 6. Temperature dependence of length at constant force for X^{20} -poly(VPGVG). Note that steep contraction occurs between 20° and 40°C and then the curves remain relatively constant with temperature. (Adapted with permission from reference 40).

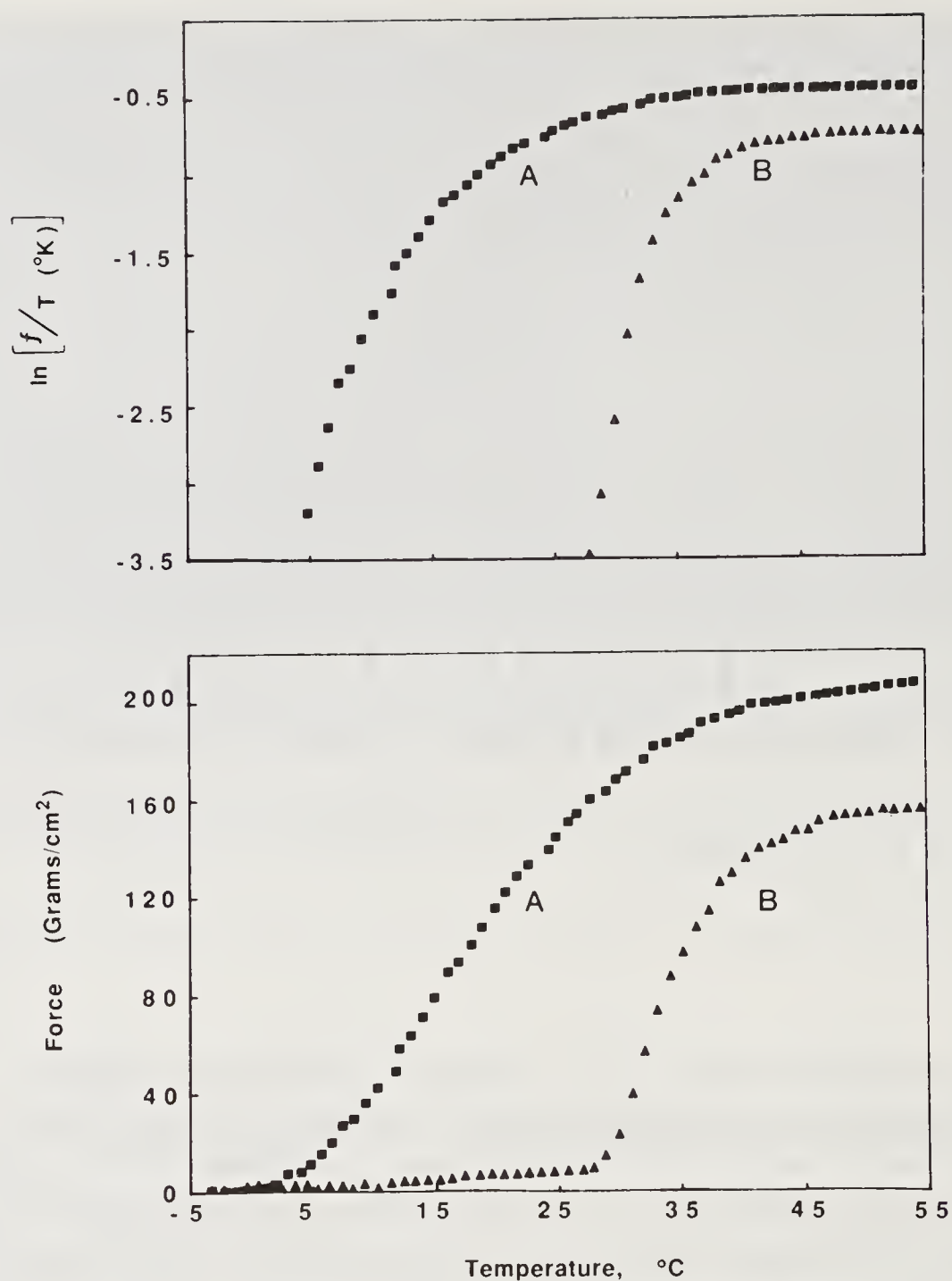


Figure 7. Temperature dependence of force, f , at constant length for X^{20} -poly(VPGVG) in ethylene glycol/ H_2O (30:70) curve A and in water curve B. The upper figure is a classic thermoelasticity curve plotted as $\ln(f/T)$ vs. T in order to determine the f_e/f ratio. See text for discussion. (Reproduced with permission from reference 41).

experimental data on both the uncross-linked and cross-linked states that elucidate the molecular processes underlying the transition. The higher temperature process will be discussed in terms of the classical argument for delineating the relative magnitudes of the internal energy and entropy terms which contribute to the total elastomeric force.

1. *Lower temperature process: The inverse temperature transition of folding and assembly:* The phenomenon of an inverse temperature transition arises from hydrophobic interactions (42–46) wherein apolar (hydrophobic) moieties such as the isopropyl side chain of the valyl residue are surrounded by waters of hydrophobic hydration which are more ordered than bulk water. The water of hydrophobic hydration has a greater heat capacity than bulk water and its formation from bulk water is exothermic (47). On raising the temperature the waters of hydrophobic hydration destructure to become less-ordered bulk water as the polypeptide becomes more-ordered with the formation of intramolecular hydrophobic contacts as in the case of folding to form the β -spiral and with the formation of intermolecular

hydrophobic contacts as when several β -spirals supercoil to form twisted filaments. This may be called thermally driven polypeptide structure development.

a. Review of experimental data describing thermally driven structure development.

In order to arrive at a true description of the nature of elasticity in these elastic-protein-based polymers, it is essential to have as clear an understanding as possible of the molecular state following the transition which is so apparent in the $(\partial l/\partial T)_f$ and the $(\partial f/\partial T)_l$ experiments (see figures 6 and 7). Because of the importance of this issue, many studies have been undertaken to achieve the most accurate molecular description.

(1) Microscopy showing thermally driven assembly into non-random anisotropic structures: The most unambiguous demonstration that these repeat sequences from elastin increase order on increasing temperature is given by the thermally driven reversible crystallization shown in figure 1. As discussed above, this argument continues to the pentapeptide repeat with the crystal structure of the cyclopentadecapeptide (see figure 2) and with the thermally driven self-assembly into anisotropic fibers demonstrated by the polypentapeptide (see figure 3). The evidence is clear that there is an increase in intermolecular order on raising the temperature.

The temperature dependence of aggregation of dilute solutions is given in figure 8a (48). When a drop of a cloudy solution is placed on a carbon coated grid, when the excess water is wicked off and when the grid is negatively stained with an oxalic acid, uranyl acetate solution at pH 6.3, the molecules of poly(VPGVG) are seen to have assembled in the thermally driven aggregation process into parallel aligned filaments with a twisted filament substructure. Optical diffraction of such micrographs gives rise to a series of diffraction spots as shown in the insert of figure 3d (23). The major equatorial deflection spot is near 50 Å indicating the diameter of the parallel aligned filaments and an off-equatorial diffraction is seen at E indicating a helical substructure to the filament. The twisted filaments of figure 4e are an effort to develop the twisted filaments of the transmission electron micrograph from the β -spiral structure (37).

Clearly the basis for an underlying filamentous, anisotropic structure is well-established starting from the fibers seen in the light microscope without fixative or staining of any kind (see figure 3a), progressing through the parallel aligned fibrils of which the fibers are comprised as seen in the scanning electron microscope (see figures 3b and 3c), and arriving at the underlying source of the anisotropic fibers, the parallel aligned filaments seen in the transmission electron microscope with negative staining and optical diffraction of figure 3d. Finally, it is the β -spirals of figures 4c and 4d which associate and supercoil to form the twisted filament that is the underlying source of the anisotropic fiber formation, and it is the details of the β -spiral that are discussed below.

(2) Nuclear magnetic resonance conformational studies: From the very first proton NMR conformational studies (27–29), the β -turn was shown to be the dominant conformational feature in VPGVG, its oligomers, and high polymers (see figure 4a). Proton NMR was then used to show the nearly identical conformations of cyclo(VPGVG)₃ and of poly(VPGVG) (34). This conformational relationship was referred to as a cyclic conformation with a linear conformational correlate (31). Detailed proton NMR studies to obtain secondary structure and torsion angle information on cyclo(VPGVG)₃ was combined with computational efforts to arrive at a conformation which was very similar to that in the crystal (see figure 9) (32).

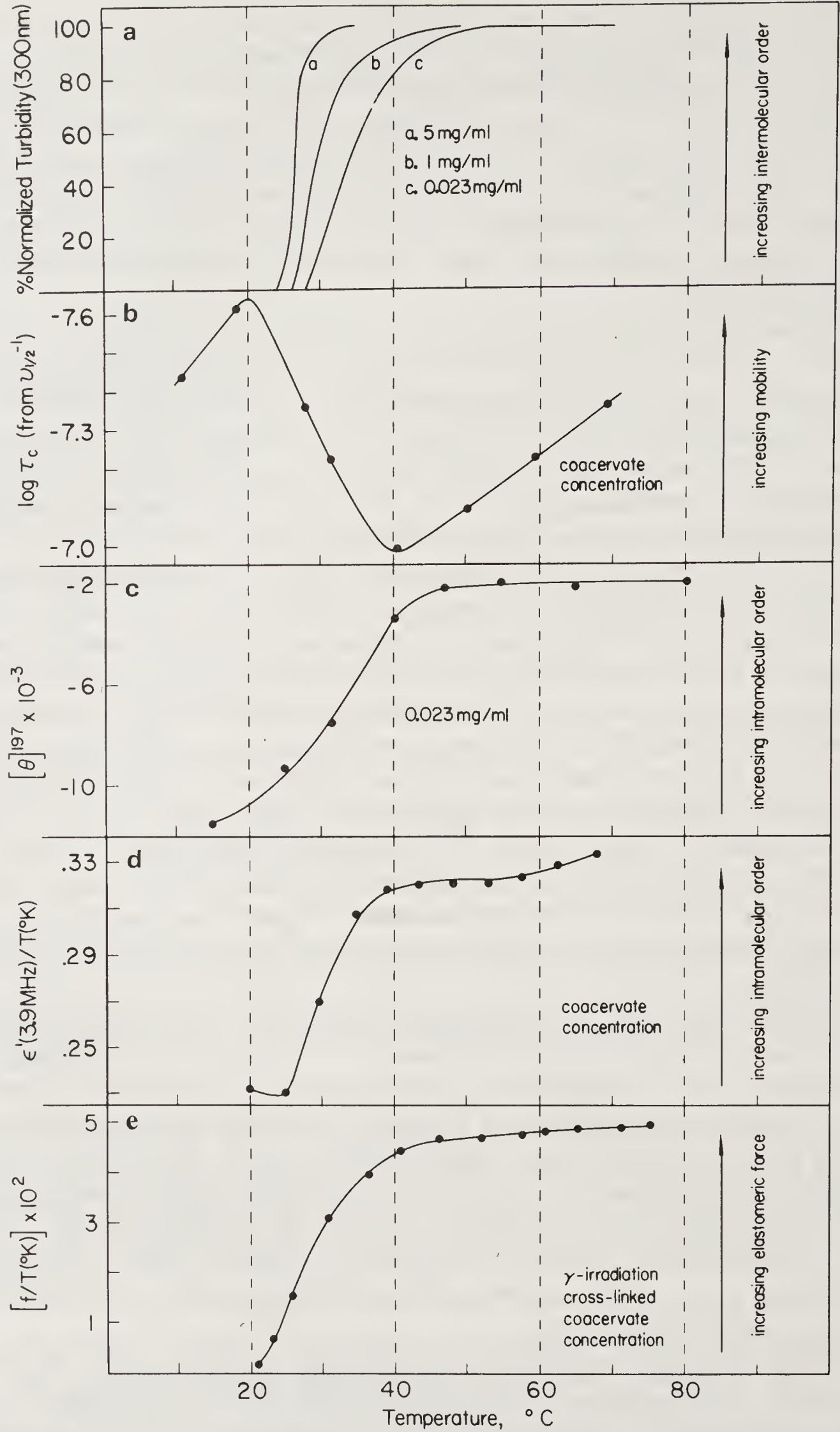


Figure 8. Comparison of five different physical characterizations of poly(VPGVG) in (a), (b), (c) and (d), with X²⁰-poly(VPGVG) in (e). See text for discussion. (Reproduced with permission from reference 48).

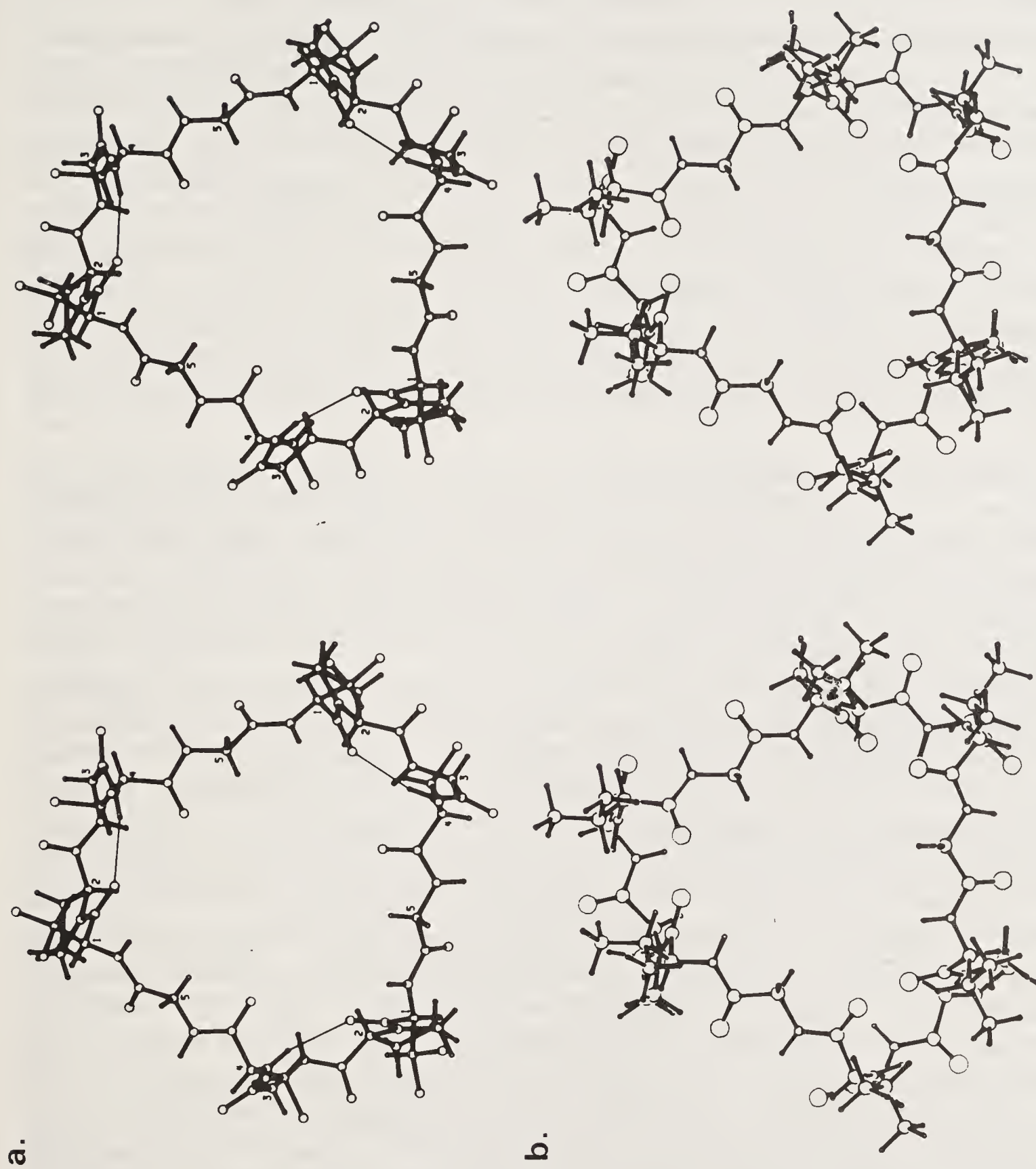


Figure 9. Molecular structure of cyclo(VPGVG)₃ in the crystalline state, (a) and as derived in solution using NMR and computations, (b) (a reproduced with permission from reference 19 and b from reference 32).

Thus, the evidence for the recurring β -turn in poly(VPGVG) was substantive and continuous from the earliest conformational studies.

(3) Raman scattering of crystal, solution, and coacervate states: Using the crystals of cyclo(VPGVG)₃ to define the Raman bands for the Type II Pro 2–Gly 3 β -turn, it was found that the β -turn was the conformation in solution for cyclo(VPGVG)₃ and for poly(VPGVG) and in the coacervate state for poly(VPGVG) (49).

(4) Circular dichroism demonstration of thermally driven increase in intramolecular order: At low temperatures the circular dichroism spectra of poly(VPGVG) in water indicate a significant presence of disorder with an ellipticity at 197 nm of about -1.2×10^4 . On raising the temperature through the temperature of the inverse temperature transition the circular dichroism spectra changes dramatically to -0.2×10^4 at 197 nm (see figure 8c) and exposes a pattern indicative of a Type II β -turn (48, 50). Interestingly, the change in ellipticity at 197 nm for a 2.3 mg/ml concentration as a function of temperature almost exactly overlies the plot of change in elastomeric force with increasing temperature (50). Even the more dilute solution of 0.023 mg/ml shows the correspondence of increase in force with increase in intramolecular order demonstrated by the ellipticity data, i.e., compare figures 8c and 8e.

(5) Nuclear magnetic resonance relaxation studies: Relaxation studies have been carried out on poly(VPGVG) in the coacervate state and after 20 mrad cross-linking to form an elastomeric cylinder. Carbon-13 and nitrogen-15 NMR studies were carried out on carbonyl carbon and peptide nitrogen enriched samples and the temperature dependence of the correlation time was determined by means of longitudinal relaxation times, transverse relaxation times and nuclear Overhauser effect magnitudes (38, 39). As the temperature was raised from well below the transition, the correlation time became shorter due to increased motion, then, as the temperature began to pass through the temperature range of the inverse temperature transition beginning at 25°C, the correlation time became longer due to decreasing mobility until the transition was complete near 37°C. As the temperature was raised above 37°C the correlation time again began to get shorter with increases in temperature. This is shown in figure 8b. Thus the inverse temperature transition became characterized within the elastomeric state itself as a decrease in mobility with increase in temperature. This occurs even when the study is carried out without a change in composition, i.e., without a change in water content, where the coacervate concentration of approximately 40% peptide and 60% water by weight is the composition chosen at the low temperature. The temperature dependence of the correlation time above 37°C indicates an activation energy for a mobility of about 1.2 kcal/mole for the peptide moieties. The relaxation was treated as being comprised of a single correlation time rather than a distribution of correlation times. A distribution of correlation times had been assumed by Lyerla and Torchia (51) in their studies on elastin on the basis of the argument that a random chain network must by necessity exhibit a distribution of correlation times as the same peptide moieties in different repeats would necessarily be in different configurational states. Significantly, the temperature range for the transition observed in the NMR relaxation studies correlates with the temperature range for the increase in elastomeric force (compare figures 8b and 8e).

(6) Dielectric relaxation studies on the coacervate composition (52): At temperatures below that of the inverse temperature transition the real part of the dielectric

permittivity shows a monotonically increasing curve as the frequency is decreased from 1 GHz to 1 MHz for the coacervate concentration. As the temperature is raised through the transition, an intense relaxation develops near 10 MHz. This relaxation is well-fit by a simple Debye expression or when fit by a Cole–Cole function shows essentially no dispersion of frequencies (6). It is not characterized by a distribution of correlation times. If, as argued by Lyerla and Torchia (51), a random chain network would be characterized by a distribution of correlation times, then this result provides a significant statement that the elastomeric state above the temperature of the inverse temperature transition is not describable as a random chain network. Considering the β -spiral of figure 4d, the net dipole moment change of a pentamer, as it undergoes its collective peptide rocking motions within a 1.5 kcal/mole cut-off energy, calculated to give the appropriate magnitude of the dielectric decrement for poly(VPGVG) and for a damped analog with a D-Ala residue in position five (53). Additionally, the magnitude of the activation energy for the rocking dipole moment of about 1.2 kcal/mole (54) is essentially the same as that found in the NMR relaxation studies (38). Furthermore, the magnitude of the correlation times obtained from dielectric relaxation and NMR relaxation studies are in good agreement (38, 39, 52, 54). Finally, the curve for the change in dielectric permittivity (real part, 3.9 MHz) as a function of temperature approximates the curve for the development of elastomeric force (compare figures 8d and 8e; also see figure 10 of reference 6).

(7) NMR nuclear Overhauser effect for through space proton–proton contacts: Two-dimensional correlation spectroscopy (2D-COSY) and two-dimensional nuclear Overhauser spectroscopy (2D-NOESY) studies have been carried out, as noted above, on cyclo(VPGVG)₃ and on poly(VPGVG) in dimethyl sulfoxide and in water at a temperature just below the onset of aggregation for poly(VPGVG) (34, 35). The 2D-NOESY maps of both were essentially the same, except for the higher resolution in the cyclic molecule, indicating again the nearly identical conformations (31). There was observed for poly(VPGVG), however, an additional proton–proton contact which occurred between Val¹ γ CH₃ protons and the Pro² β CH₂ protons. As this direct contact is not possible in the Val 1–Pro 2 sequence, it requires that the Val¹ γ CH₃ and Pro² β CH₂ protons come from different pentamers. Since it was not present in the cyclic molecule but was observed in the linear high polymer, this proton–proton contact is a candidate for intramolecular interturn interactions. In this case it would be a direct demonstration of the intramolecular hydrophobic contacts arising out of the inverse temperature transition. Furthermore, depending on whether the contact involved the Val¹ γ CH₃ or the Val⁴ γ CH₃ protons, an approximate number of pentamers per turn of β -spiral could be deduced. In the initial computational generation of the β -spiral (33), a number of 2.7 pentamers per turn was calculated which would require that the Val⁴ γ CH₃ provide the γ -protons for the interaction. The original 2D NOESY data (34), however, suggested that the Val¹ γ CH₃ moiety provided the γ -protons for interaction with the Pro² β CH₂ protons. In order to determine which residue was responsible, Val¹(d₈)-poly(VPGVG) and Val⁴(d₈)-poly(VPGVG) were synthesized. Using these polymers, it was found the 80% or more of the interaction was due to the Val¹ γ CH₃ protons interacting with the Pro² β CH₂ protons. This indicates that a β -spiral with approximately three pentamers/turn was the dominant state and a calculated low energy structure giving 2.9 pentamers per turn was reported (35). Interestingly, Wasserman and Salemme using a different molecular dynamics approach and with poly(VPGVG) immersed in water obtained exactly three

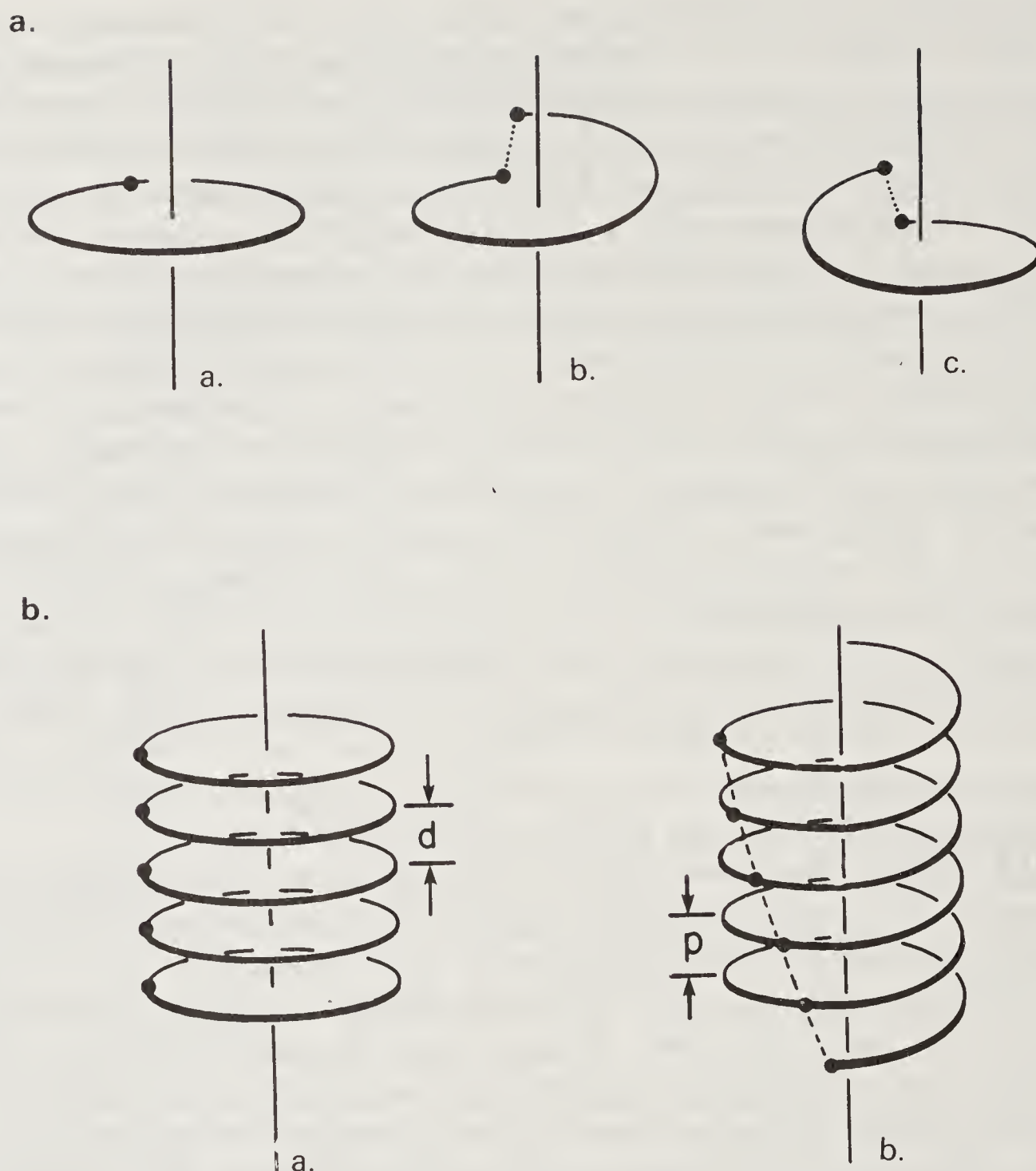


Figure 10. Schematic representation of the concept of a cyclic conformation and the introduction of a helical pitch. (a) To result in a linear helical conformational correlate. (b) See text for discussion. (Reproduced with permission from reference 33).

pentamers/turn (36). Of course, consistent with all of the above reviewed data, this requires that poly(VPGVG) be a non-random structure.

b. Generation of the β -spiral of poly(VPGVG) from the cyclic conformational correlate of cyclo(VPGVG)₃.

When there are some 15 residues in a cyclic structure there would be only small torsion angle changes to give the 15 residues a pitch suitable for the formation of a helical structure. This process is indicated in figure 10a to achieve either a right-handed (b) or left-handed (c) helical sense. It then becomes apparent that a stack of such cyclic structures could readily be converted to a helical structure as shown in figure 10b for conversion to a right-handed helix with only a small perturbation of the cyclic structure. This is known as the concept of cyclic conformations with linear conformational correlates and a stack of cyclo(VPGVG)₃ molecules in the crystal structure in figure 2 would be the cyclic correlate of the linear helix. This concept was first enunciated as one of three approaches in deriving the conformation of the linear alternating L-D polydipeptide of Gramicidin A from the cyclo(LDLDLD) structure of enniatin B and in deriving related structures (55); it was immediately used

conceptually to consider β -spiral structures of the repeat peptides of elastin (27); it was used experimentally with a comparison of the cyclic series cyclo(VPGVG)_n with $n = 1, 2, 3, 4, 5$ and 6 to determine which, if any, cyclic moieties had conformations similar to poly(VPGVG) (31); and it was used computationally to generate the linear conformational correlate of cyclo(VPGVG)₃ (33) which had been shown to be of a nearly identical conformation to poly(VPGVG) on the basis of extensive proton and carbon-13 NMR studies (31).

Specifically, the NMR and computationally derived solution conformation of cyclo(VPGVG)₃ (32) was cleaved to give a linear pentadecapeptide and thereafter given a right- and then a left-handed helical pitch using the computational approach of Go and Okuyama (56) and Go and Scheraga (57, 58). Only when the helix with the right-handed helical sense was generated was a linear conformational correlate obtained to give a class of dynamic conformations one of which is represented in figure 4d. It was the obtaining of this class of dynamic helices, called β -spirals, in which the β -turn functioned as spacers between the turn of the helix and in which there were the Val 4–Gly 5–Val 1 segments suspended between β -turns that provided the insight (14) to arrive at the basis for entropic elastomeric force in a non-random, regular and dynamic helical structure. The new mechanism was called the librational entropy mechanism of elasticity (37) which is the damping of internal chain dynamics on extension. The proposed structure and new mechanism suggested a series of experimental and computational approaches by which they could be tested, and these will be briefly reviewed below. First, however, will be a short presentation of the mechanical studies that determine the nature of the elasticity.

2. *Higher temperature process: demonstration of a dominantly entropic elastomer:* The total elastomeric force, f , exhibited by an elastomer is the sum of internal energy, f_e , and entropy, f_s , components, i.e.,

$$f = f_e + f_s. \quad (1)$$

The classical approach to determine the relative magnitudes of f_e and f_s is due to Flory *et al.* (59) wherein the f_e/f ratio is obtained from the slope when $\ln(f/T)$ is plotted vs. T on the basis of equation 2, i.e.,

$$\frac{f_e}{f} = -T \left(\frac{\partial \ln(f/T)}{\partial T} \right)_{V,L,n}, \quad (2)$$

where the experiment is carried out with the elastomer extended to a fixed length, L , and at constant volume, V , and composition, n . Commonly, a correction term is required for the conditions of constant pressure, P , and at equilibrium, *eq*, with a surrounding solvent. In this case the expression becomes (60).

$$\frac{f_e}{f} = -T \left(\frac{\partial \ln(f/T)}{\partial T} \right)_{P,L,eq} - \frac{\beta_{eq}T}{\alpha^3(V_i/V) - 1}, \quad (3)$$

where $\beta_{eq} = (\partial \ln V / \partial T)_{P,L,eq}$ is the thermal expansion coefficient; α is the fractional increase in length L/L_i where L_i is the initial length and L the length at fixed extension, and V_i and V are the elastomer volumes at length L_i and after elongation to length L . With a determination of all the necessary values in equation 3 for a cylinder of X²⁰-poly(VPGVG) (39) and arising from a detailed composition study (61), the f_e/f

ratio was determined to be 0.12 in the temperature interval of 40° to 65°C (6). Because the composition and volume of the non-extended state are essentially constant in the 40° to 60°C temperature range (61), it is possible to use equation 2 in relation to the data in the upper part of figure 7. It is apparent from figure 7 that, once the lower temperature process of the inverse temperature transition is complete, the higher temperature process exhibits a near zero slope in the $\ln(f/T)$ vs. T plot indicative of a dominantly entropic elastomer. Specifically, the f_e/f ratio is of the order of 0.1 or less for X²⁰-poly(VPGVG) in water (curve B) as well as for X²⁰-poly(VPGVG) in ethylene-glycol water (30:70) (curve A).

These results require the conclusion that X²⁰-poly(VPGVG) is a non-random entropic elastomer and necessitate a discussion of the mechanism of elastomeric polypeptides.

Source of Entropic Elastomeric Force in Non-Random (Self-Assembling) Molecular Systems

There have been three mechanisms put forward for the source of entropic elastomeric force in polymers: (1) the classical theory of rubber elasticity based on random chain networks (62–66), (2) a hydrophobic hydration mechanism arising out of the ordering of solvent around hydrophobic side chains which become exposed on stretching (67–70), and (3) a librational entropy mechanism wherein there occurs a damping of internal chain dynamics on extension (14, 37, 71–74). Both (1) and (3) depend on changes in chain configuration entropy but each arises out of fundamentally different concepts of the relevant molecular structure whereas (2) is based on changes in solvent entropy.

Random chain network perspective: In the classical theory of rubber elasticity, the Eyring equation (75) for the mean square end-to-end chain length, $\langle r^2 \rangle_0$ is the starting point for the calculation of f_e/f , i.e.,

$$\frac{f_e}{f} = T(d \ln \langle r^2 \rangle_0 / dT) \quad (4)$$

The calculation is carried out for a single chain *in vacuo* and initially allows a chain to pass through itself (a phantom chain) in deriving the distribution of end-to-end chain lengths. In doing so it samples the rotation about each virtual bond or actual bond of the backbone. This approach has been applied to poly(VPGG), poly(VPGVG) and poly(APGVGV) (76, 77) describing each as a dominantly entropic elastomer which, when the β -turn is included, results in the case of poly(VPGVG) with a calculated f_e/f ratio of 0.25. This is essentially that found experimentally (0.26) for bovine elastin (78) and represents a satisfying finding for the random chain network approach. Interestingly, the calculation also describes poly(APGVGV) as a dominantly entropic elastomer with an even smaller f_e/f ratio of 0.08, i.e., a more dominant entropic component, and as a soft material with low-force development on extension. However, the polyhexapeptide, poly(APGVGV), was earlier described as a non-elastomeric polypeptide (79) and when mixed with poly(VPGVG) and γ -irradiation cross-linked functions to make the matrix more stiff (80) in analogy to a hard segment of polyurethane. Accordingly, the random chain network gives an incorrect characterization of poly(APGVGV); presumably this is because it is

experimentally found to be a rigid molecule with substantial secondary structure (15). Also the extensive physical characterization of poly(VPGVG) argues that it is not well-described as a random chain network; it is indeed dynamic but with the regular helical structure arising due to optimization of non-restricting intramolecular interturn hydrophobic contacts rather than arising due to restricting hydrogen bonding between repeats. Additionally, when computing distributions of end-to-end lengths, instead of changing a single torsion angle which would cause a dramatic change in the position of one end with respect to the other end, if a second torsion angle were varied, as occurs with the rocking of a peptide moiety, then there would be little change in the position of one end with respect to the other. In this case the entropy gain due to the torsion angle changes would be retained while maintaining a similar end-to-end chain length. This is what is proposed in the librational entropy mechanism.

Solvent entropy mechanism: It is appreciated that when an apolar (hydrophobic) moiety is exposed to water, there is a change in the entropy of the water on hydration of the hydrophobic moiety. The question is whether this solvent entropy change actually contributes to the entropic elastomeric force. One way to approach the question is to use solvent mixtures and to follow the heat required in the phase transition of coacervation. During coacervation the more-ordered water surrounding the hydrophobic moieties becomes less-ordered bulk water. For such a phase transition the heat of the transition divided by the temperature of the transition gives the entropy for the transition. Thus in the differential scanning calorimetry data of figure 11, it is seen that, as the solvent composition is changed by the addition of ethylene glycol (EG), the heat of the transition decreases with only a small decrease in temperature of the transition such that the entropy of the transition is dramatically decreased by a 30% ethylene glycol/70% water solvent mixture. Now, as stretching of the elastomer exposes the hydrophobic side chains to solvent in a reversal of the dominant process of the phase transition, such an exposure of hydrophobic groups in the EG:H₂O (30:70) solvent mixture would lead to a decrease in change in solvent entropy. This follows the approach used by Hovee and Flory (81, 82) for elastin. The prediction would be that, if solvent entropy contributes to elastomeric force, the decrease in solvent entropy change in an ethylene glycol–water mixture would result in a decreased elastomeric force exhibited by the elastomer. In fact the reverse occurs. The elastic modulus of X²⁰-poly(VPGVG) is greater in EG:H₂O (30:70) than in water (41). This is apparent in the thermoelasticity data of figure 7 where curve A is in EG:H₂O (30:70) and curve B is in pure water. Thus the experimental data indicate that changes in solvent entropy do not contribute in a definitive way to the entropy component of elastomeric force. In both solvents the data indicate dominantly entropic elastomeric force with f_e/f ratios of 0.1 or smaller.

Librational entropy mechanism: The librational entropy mechanism arises out of the class of dynamic β -spiral conformations derived for poly(VPGVG) as depicted in figure 4. On deriving this family of conformations, it was immediately apparent that the peptides in the Val 4–Gly 5–Val 1 suspended segment ought to be able to undergo large amplitudes rocking motions and that the damping of these large torsional oscillations, or librations, on extension could be the source of entropy decrease responsible for elastic resistance to extension and for the restoring force on removal of the extending force. The perspective also immediately suggested a number

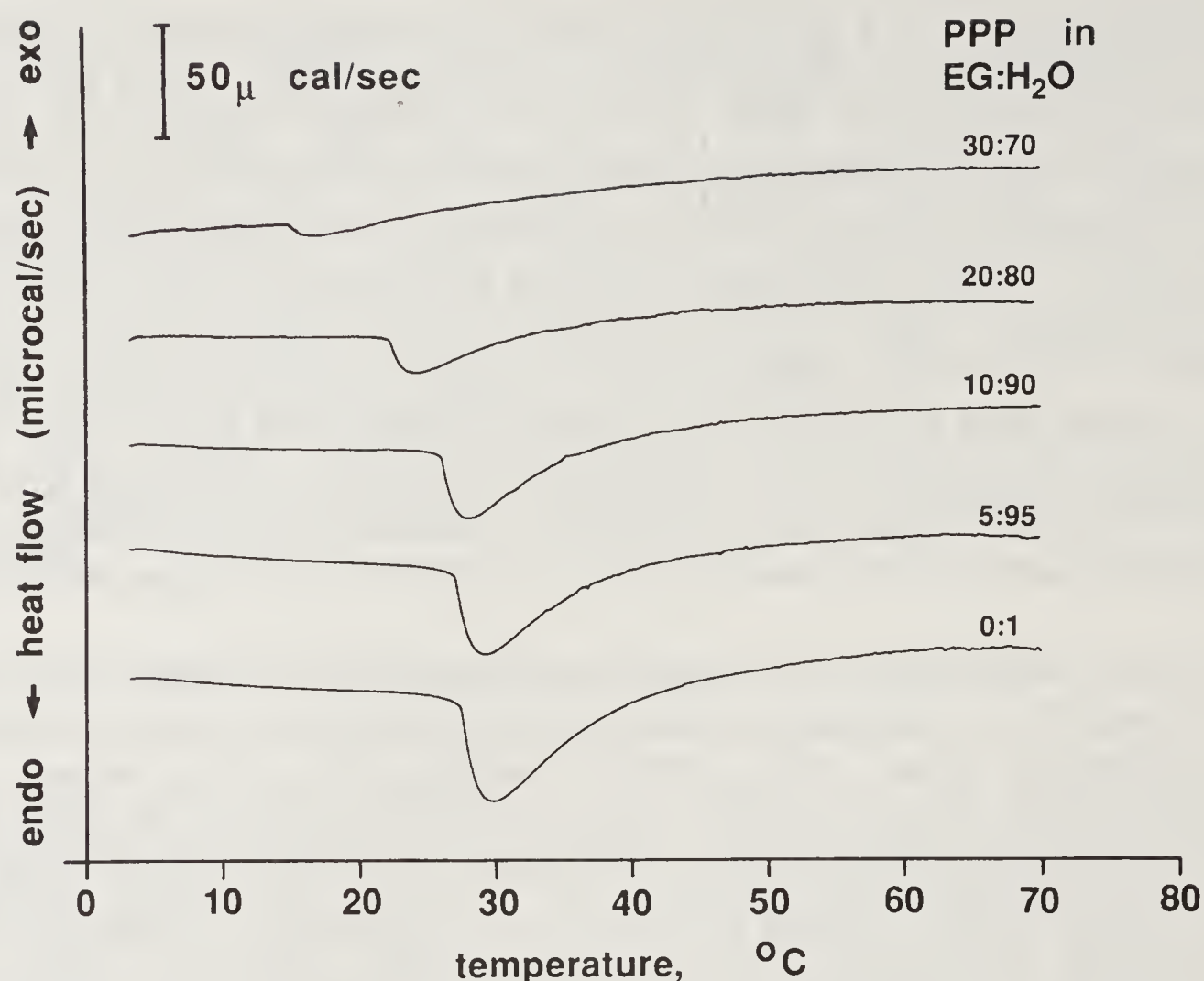


Figure 11. Differential scanning calorimetry data for poly(VPGVG), indicated as PPP, as a function of increasing ethylene glycol(EG) in the solvent. The endothermic heat of the transition, due to the thermal destructuring of waters of hydrophobic hydration, is reduced to near zero at a 30:70, EG:H₂O solvent mixture. At this stage the entropy change on solvation and desolvation is very small. (Reproduced with permission from reference 41).

of experimental and computational approaches to test the hypothesis. Experimentally: (1) synthesize the analogs L-Ala 5 and D-Ala 5 to see if the introduction of side chains at position 5 would have an effect on elastic properties by decreasing amplitudes of libration and (2) carry out dielectric relaxation, as well as NMR relaxation, studies to see if a relaxation process could be defined which could be assigned to the rocking of peptide moieties. Computationally: (1) calculate the potential energy contours for the peptide rocking process to see if there is a correlation between changes in the ϕ_{i+1} torsion angle and the ψ_i torsion angle and to see if there is a decrease in the number of states allowed for a given cut-off energy on extension and (2) calculate the molecular dynamics to see if there is a greater amplitude of libration of the peptides in the Val 4–Gly 5–Val 1 segment and to see if the amplitude of such librations become damped on extension.

a. Experimental tests of proposed librational entropy mechanism

(1) Synthetic blocking of motional freedom in suspended segment: The large amplitude motions suggested to occur in the two peptide moieties in the Val 4–Gly 5–Val 1 segment are possible in part due to the absence of a side chain for residue 5. The synthetic approach therefore was to add the minimal sized side chain with the synthesis of (L-Val 1–L-Pro 2–Gly 3–L-Val 4–L-Ala 5)_n, i.e., poly(VPGVA), and of (L-Val 1–L-Pro 2–Gly 3–L-Val 4–D-Ala 5) designated as poly(VPGVA'). Synthesis of poly(VPGVA) resulted in a polypeptide with retention of the recurring β -turn and with essentially the same temperature profile for aggregation (perhaps raised 1° to

2°C for the same concentration) (83). The result of raising the temperature of aqueous solutions of poly(VPGVA), however, is not the formation of a viscoelastic coacervate; rather it is the formation of a granular precipitate. Efforts to γ -irradiation cross-link the granular precipitate give rise to a sheet of material that simply fragments when stress/strain studies are attempted (83). This addition of a CH_2 moiety is in stark contrast to the addition of a CH_2 moiety to residue 1, i.e., poly(IPGVG), where I = Ile in which the temperature of the transition is lowered by 20°C with the conformation being identical with poly(VPGVG) before and after the transition and with an elastomeric γ -irradiation cross-linked product having properties very similar to X^{20} -poly(VPGVG) (84, 85).

Synthesis of poly(VPGVA') resulted in a polypentapeptide which again retained the recurring β -turn and which exhibited a temperature profile for aggregation that was lowered by 3° or 4°C. The product of the thermally driven aggregation was a slowly forming coacervate but the γ -irradiation cross-linked product did not give a functional elastomer, rather a band was obtained that simply disintegrated on efforts to stretch in water (86).

(2) Dielectric relaxation studies of the polypentapeptide and analogs: As mentioned previously the dielectric relaxation of poly(VPGVG) coacervate concentration on going from 20° to 40°C resulted in the development of an intense relaxation in the 10 MHz range (52); the 40°C minus 20°C difference curve is plotted in figure 12b for the real part of the dielectric permittivity where the experimental curve is seen to be well fit by a Debye expression with a $\Delta\epsilon'$ of 72. This relaxation is assigned to a peptide librational mode (52). Dielectric relaxation study of poly(VPGVA) gave no results whereas that for poly(VPGVA'), i.e., the D-Ala 5 polypentapeptide, did give rise to a similar but damped relaxation as shown in figure 12a (R. Henze and D. W. Urry, unpublished data). Interestingly, computation of the $\Delta\epsilon'$ from the β -spiral structures with 0.6, 1.0, 1.5, and 2.0 kcal/mole pentamer cut-off energies, using the Onsager equation for polar liquids (87) to relate a calculated dipole moment change to the experimental $\Delta\epsilon'$, gave a calculated dipole moment change within a factor of two of the experimentally derived dipole moment change and more significantly gave the correct ratio for the observed damping of figure 12 (53). Both the experimental finding of a peptide librational mode and the capacity to calculate the magnitude of the dielectric decrement, $\Delta\epsilon'$; and the extent of damping due to the D-Ala 5 residue provided satisfying support for the peptide librational process. It is now of interest to estimate the contribution of the expected damping due to extension on a decrease in entropy required for an entropic component of the elastomeric force.

b. Computational tests of the proposed librational entropy mechanism:

(1) Molecular mechanics estimates of changes of volumes in configuration space on extension: Starting with the β -spiral conformation, the question is asked as to the number of ways a pentamer can span a pentamer-sized gap in the helix within a given energy cut-off above the lowest energy conformation. Each 5° rotation of a torsion angle is counted as an additional state. For 0.6, 1.0, and 2.0 kcal/mole-residue cut-off energies, the number of states for the relaxed conformation are 342, 762, and 1853, respectively, and on extension to 130% the number of states become 24, 58 and 162, respectively (71, 72, 53). Using the Boltzmann relation for the entropy of a mole of pentamers

$$S = R \ln W, \quad (5)$$

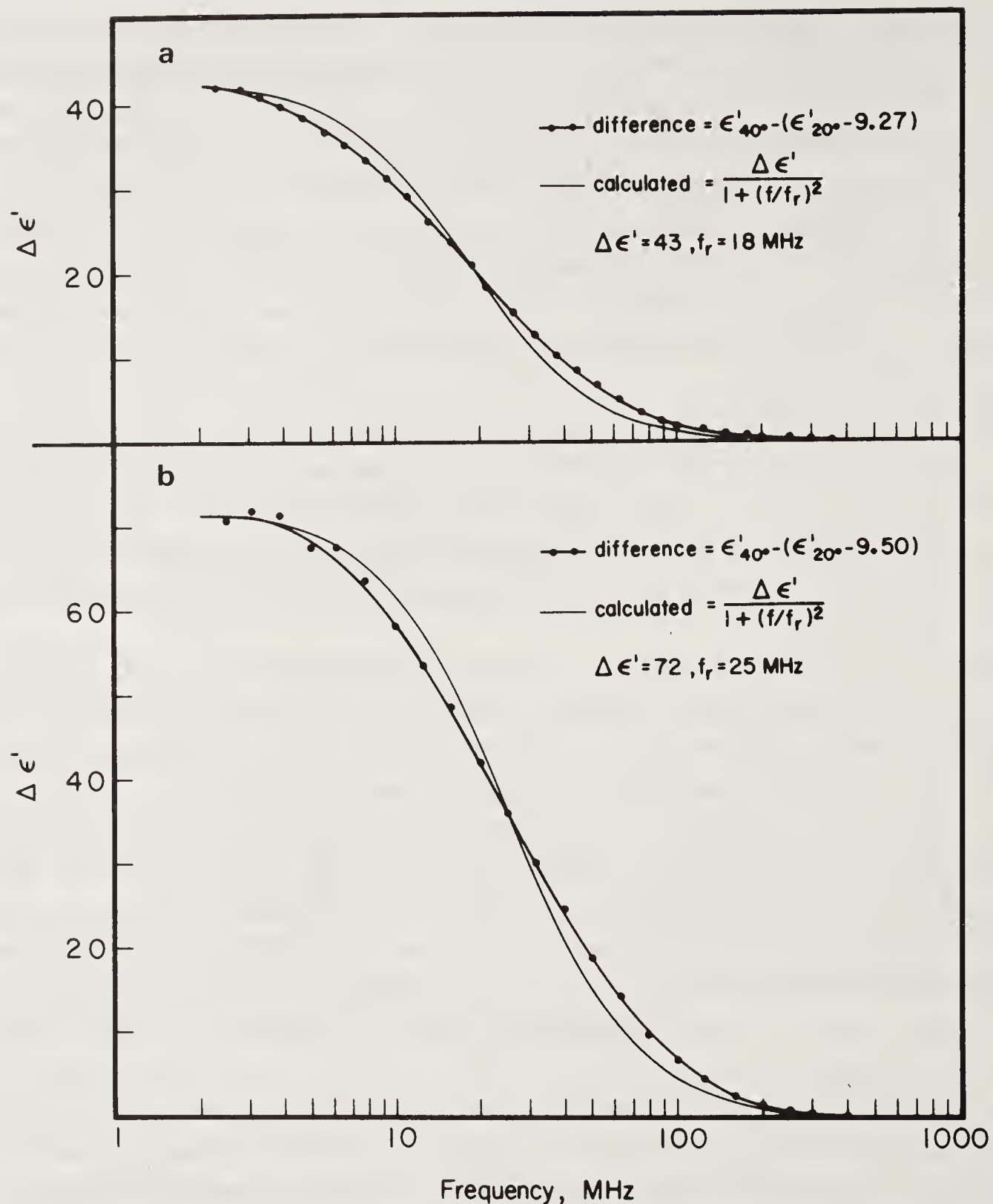


Figure 12. Dielectric relaxation data for (a) Poly(L-Val 1-L-Pro 2-Gly 3-L-Val 4-D-Ala 5) designated poly(VPGVA') and (b) Poly(VPGVG). The curve through the data points is closely approximated by fitting to a simple Debye expression showing an intense relaxation with a $\Delta\epsilon'$ of 72 for poly(VPGVG) but damped to 43 for poly(VPGVA'). (a. Henze and Urry, unpublished data; b reproduced with permission from reference 6).

where W is the number of *a priori* equally probable states accessible to the pentamer in the β -spiral and R is the gas constant 1.87 cal/mole-deg, and writing the entropy change, ΔS , on going, by stretching, from the relaxed, r , to the extended, e , states to be

$$\Delta S = S^r - S^e = R \ln \frac{W^r}{W^e}, \quad (6)$$

gives the values in table 1. More accurately the states can each be properly weighted for its energy by means of a Boltzmann sum over states, and the statistical mechanical

Table 1. Perspective of entropy of the polypentapeptide β -spiral by enumeration of states

Cutoff energy	Number of States		Entropy change per residue
	Relaxed (r)	Extended (e)	
2 kcal/mole	1853	162	0.97
1 kcal/mole	762	58	1.02
0.6 kcal/mole	342	24	1.06
using Boltzmann sum over states			
$f = \sum_i e^{-\varepsilon_i/kT}$			
2 kcal/mole	$\Delta S = R \ln(f^r/f^e) + \frac{E^r - E^e}{T}$		1.01

definition of entropy, i.e.,

$$\Delta S = R \ln(f^r/f^e) + \frac{E^r - E^e}{T}, \quad (7)$$

where the partition function, f , is the Boltzmann sum overstates,

$$f = \sum_i e^{-\varepsilon_i/kT},$$

with ε_i being the energy of the state and k the Boltzmann's constant where $Nk = R$ with N being Avogadro's number. Since $E^r - E^e$ is small for the entropic elastomer, the second term in equation 7 is neglected. The result in each case is one entropy unit per residue for a 130% extension. This is a very substantial entropy change, quite enough to provide the entropic restoring force.

When the values of ψ_i and ϕ_{i+1} are plotted for the peptide moieties of the suspended segment, the lambda plots of figure 13 are obtained. It is immediately seen that the angles are anti-correlated, that is, the data points for each state fall approximately on a line at 45° with axes such that a positive change $\phi(\text{Gly } 5)$ for example is compensated for by a negative change of similar magnitude in $\psi(\text{Val } 4)$ (see figure 13a). This demonstrates the peptide librational process. Interestingly, the number of states and allowed excursions are dramatically reduced on extension of the poly(VPGVG) β -spiral as seen in figure 13b. Furthermore, the addition of a CH_2 moiety at residue 5, giving the D-Ala 5 polypentapeptide also exhibits the damped amplitude. This restriction of motion is even greater in the lambda plot for poly(VPGVA) (71). Thus the molecular mechanics approach substantiates the proposed librational entropy mechanism.

(2) Molecular dynamics estimates of entropy from torsion angle fluctuations: Instead of using a potential energy surface to assess the available volume in configuration space accessible to the β -spiral, it is useful to use a molecular dynamics approach and compute the entropy change on extension from the damping of root mean square (RMS) fluctuations of torsion angles, $\Delta\phi_i$ and $\Delta\psi_i$. The expression

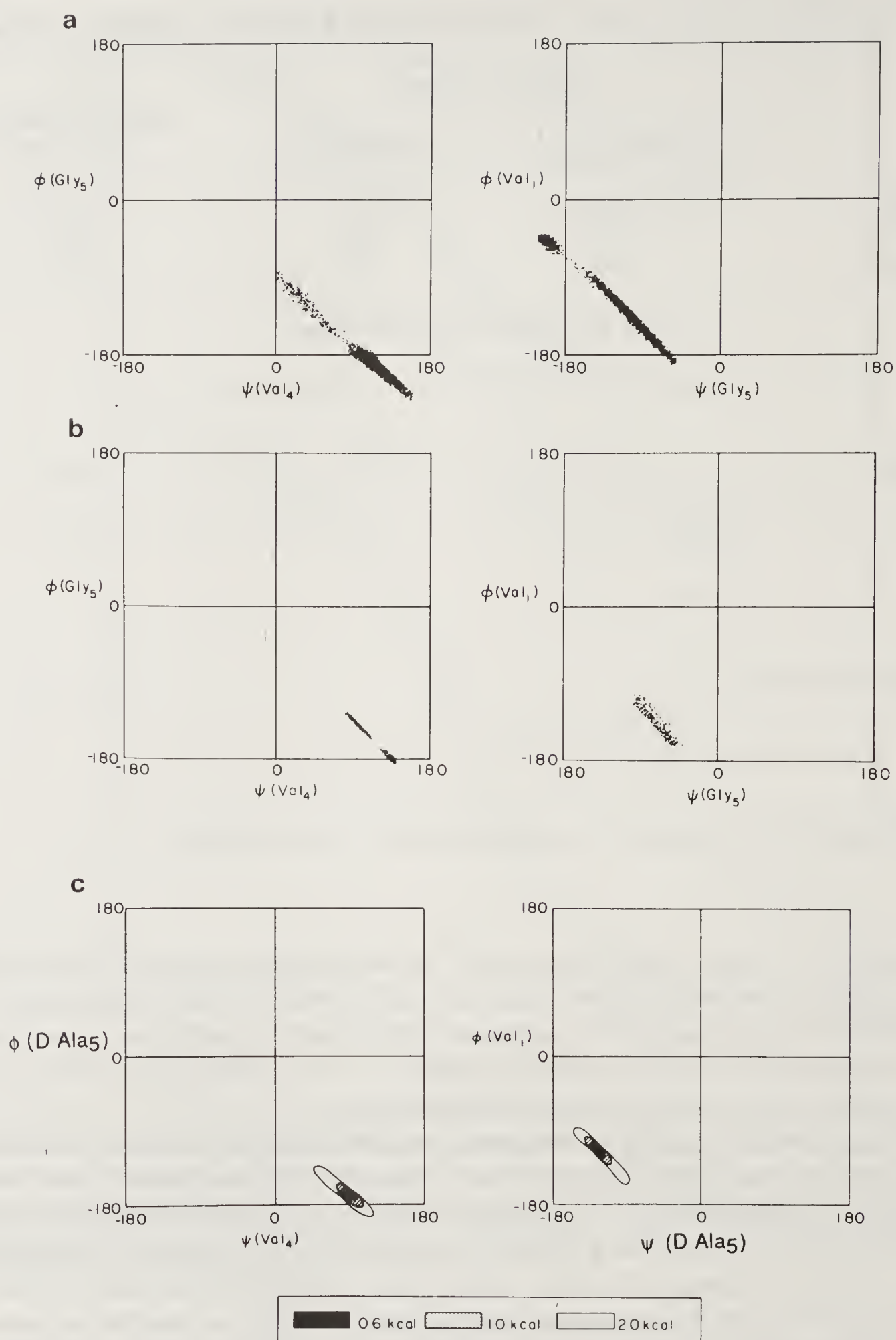


Figure 13. Lambda plots for the two peptide moieties in the suspended segment of Val 4–Gly 5–Val 1 when relaxed, (a) and when stretched 130%, (b) and in the suspended segment of L-Val 4–Gly 5–D-Ala 5 (c). See text for discussion (a and b reproduced with permission from reference 37 and c from reference 53).

may be given as

$$\Delta S = R \ln \frac{\prod_i \Delta \phi_i^e \cdot \Delta \psi_i^e}{\prod_i \Delta \phi_i^r \cdot \Delta \psi_i^r} \quad (8)$$

The CHARMM molecular dynamics approach developed by Karplus and his colleagues (88–90) and adapted by Polygen Corp. was used (74). The result of the root mean square (RMS) fluctuation of torsion angles (ϕ and ψ) of (VPGVG)₁₁ after a 45 ps equilibration time (following thermalization to 300°K) and utilizing an 80 ps molecular dynamics simulation gave the values of table 2. Removing the end effects by deleting residues one through fifteen and 46 through 55, it is seen that, in the relaxed state, there are large fluctuations in the dihedral angles of the suspended segment, as had been proposed, and that these fluctuations become dramatically damped on extension to 130%. Interestingly, the ΔS , calculated from equation 8 using the values of table 2, gives a value of -1.1 entropy units per residue (74). This is essentially the same value as obtained using the molecular mechanics approach of

Table 2. Root mean square (RMS) fluctuation of torsion angles (ϕ and ψ) of (VPGVG)₁₁(45 ps of equilibration time and 80 ps of molecular dynamics simulation.)

	angle	relaxed	extended	angle	relaxed	extended	angle	relaxed	extended
β -turns	ψ_{16}	10.87	14.17	ψ_{26}	27.33	07.64	ψ_{36}	20.19	14.84
	ϕ_{17}	09.86	15.18	ϕ_{27}	11.71	08.33	ϕ_{37}	08.99	10.47
	ψ_{17}	47.59	46.68	ψ_{27}	11.70	13.51	ψ_{37}	21.53	32.96
	ϕ_{18}	61.70	47.41	ϕ_{28}	08.61	10.36	ϕ_{38}	11.15	10.66
	ψ_{18}	09.37	16.05	ψ_{28}	09.33	08.16	ψ_{38}	11.09	27.29
	ϕ_{19}	14.25	08.67	ϕ_{29}	09.70	07.31	ϕ_{39}	12.70	10.24
suspended segments	ψ_{19}	44.09	10.99	ψ_{29}	47.32	10.48	ψ_{39}	52.00	12.50
	ϕ_{20}	41.94	09.29	ϕ_{30}	48.57	11.39	ϕ_{40}	55.88	08.37
	ψ_{20}	14.50	11.15	ψ_{30}	42.56	10.62	ψ_{40}	40.67	11.08
	ϕ_{21}	27.13	24.17	ϕ_{31}	11.43	11.38	ϕ_{41}	36.44	19.06
β -turns	ψ_{21}	09.39	22.73	ψ_{31}	12.17	09.21	ψ_{41}	12.97	14.80
	ϕ_{22}	09.94	08.00	ϕ_{32}	09.90	08.93	ϕ_{42}	11.59	07.33
	ψ_{22}	11.58	16.13	ψ_{32}	15.30	10.80	ψ_{42}	11.34	13.17
	ϕ_{23}	16.37	09.33	ϕ_{33}	09.60	07.62	ϕ_{43}	09.23	09.76
	ψ_{23}	14.33	14.25	ψ_{33}	09.88	09.43	ψ_{43}	10.60	12.53
	ϕ_{24}	11.39	29.20	ϕ_{34}	11.86	09.71	ϕ_{44}	11.06	12.82
suspended segments	ψ_{24}	19.53	37.87	ψ_{34}	63.80	08.36	ψ_{44}	41.89	35.22
	ϕ_{25}	25.02	23.06	ϕ_{35}	91.70	10.20	ϕ_{45}	48.98	31.31
	ψ_{25}	49.32	32.10	ψ_{35}	15.03	11.51	ψ_{45}	42.05	56.89
	ϕ_{26}	31.43	27.24	ϕ_{36}	21.49	18.66	ϕ_{46}	21.55	30.33

Scheraga and colleagues. (Note that the logarithmic ratio is reversed in equation 6 when compared to equation 8, thereby reversing the sign).

In a completely independent effort, and the most extensive to date, Wasserman and Salemme (36) obtained very similar results using the molecular dynamics program AMBER (91). In a lambda plot of their results they reported data quite similar to that in figure 13a for a Gly 5–Val 1 peptide moiety. Also for a 75% extension they observed the damping of internal chain dynamics on extension and obtained a decrease in entropy of about 1 cal/mole-deg. Importantly, this was carried out for a 90 residue poly(VPGVG) in a box of water molecules with some 1500 water molecules required for hydration.

Accordingly, the set of experimental and computational approaches which were designed to test the peptide librational entropy mechanism do much to substantiate this new mechanism for entropic elastomeric force in elastomeric polypeptides, and by extension it may be asked whether damping of internal chain dynamics on extension may contribute to the entropic component of elastomeric force in other elastomers.

A Tribute to G. N. Ramachandran

This new mechanism for entropic elastomeric force was first described in detail in the Festschrift volume celebrating the sixtieth birthday of G. N. Ramachandran as the article entitled *Dynamic β -spirals and a Librational Entropy Mechanism of Elasticity* (37). It is a particular pleasure, therefore, to update the development in the eight intervening years as a token of thanks and of appreciation for what this preeminent scientist has given to our scientific community. His gifts to us measure monumentally both in terms of his direct scientific contribution and in terms of the many students who have come out of the Ramachandran school. His legacy is great.

Acknowledgements

The author wishes to acknowledge the many members of the Laboratory of Molecular Biophysics, past and present, who have contributed so extensively to the work reviewed here, with particular thanks to Chi-Hao Luan and D. K. Chang for the most recent work contained in figures 7 and 11 and table 2. This work was supported in part by grant HL29578 from the National Institutes of Health and by contract N00014-89-J-1970 from the Department of the Navy, Office of Naval Research.

References

1. Sandberg, L., Leslie, J., Leach, C., Torres, V., Smith, A. & Smith, D. (1985) *Pathol. Biol.* **33**, 266–274.
2. Yeh, H., Ornstein-Goldstein, N., Indik, Z., Sheppard, P., Anderson, N., Rosenbloom, J., Cicila, G., Yoon, K. & Rosenbloom, J. (1987) *Collagen and Related Research* **7**, 235–247.
3. Indik, Z., Yeh, H., Ornstein-Goldstein, N., Sheppard, P., Anderson, N., Rosenbloom, J., Peltonen, L. & Rosenbloom, J. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5680–5684.
4. Smith, D. W., Sanberg, L. B., Leslie, B. H., Wolt, T. E., Minton, S. T., Myers, B. & Rucker, R. B. (1981) *Biochem. Biophys. Res. Commun.* **103**, 880–885.

5. Cicila, G., Yoon, K., Ornstein-Goldstein, N., Indik, Z. K., Boyd, C., May, M., Cannizzaro, L. A., Emanuel, B. S. & Rosenbloom, J. (1985) *Extracellular Matrix: Structure and Function* (Alan R. Liss, Inc., New York), 333–350.
6. Urry, D. W. (1988) *J. Protein Chem.* **7**, 1–34.
7. Urry, D. W. (1988) *J. Protein Chem.* **7**, 81–114.
8. Urry, D. W. (1990) in *Protein Folding: Deciphering the Second Half of the Genetic Code*, eds. Gierasch, L. & King, J. (American Association for the Advancement of Science), pp. 63–71.
9. Urry, D. W. (1991) in *Proteins: Structure, Dynamics and Design*, eds. Venugopalakrishnan, V., Carey, P. R., Smith, I. C. P., Huang, S. G. & Storer, A. C. (Escom Science Publishers B. V., Leiden, The Netherlands) pp. 352–360.
10. Urry, D. W., Haynes, B., Zhang, H., Harris, R. D. & Prasad, K. U. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3407–3411.
11. Urry, D. W. (1988) *Int. J. Quantum Chem., Quantum Biol. Symp.* **15**, 235–245.
12. Urry, D. W. (1990) *Mater. Res. Soc. Symp. Proc.*, Materials Research Society **174**, 243–250.
13. Urry, D. W. (in press) Federal Drug Administration/State of Maryland, Intl. Biotechnology Conf.
14. Urry, D. W. (1982) in *Method in Enzymology*, eds. Cunningham, L. W. & Frederiksen, D. W. (Academic Press, Inc., New York) **82**, pp. 673–716.
15. Urry, D. W. & Long, M. M. (1976) *CRC Crit. Rev. Biochem.* **4**, 1–45.
16. Urry, D. W. (1984) *J. Protein Chem.* **3**, 403–436.
17. Urry, D. W., Long, M. M. & Sugano, H. (1978) *J. Biol. Chem.* **253**, 6301–6302.
18. Karle, I. (private communication)
19. Cook, W. J., Einspahr, H. M., Trapane, T. L., Urry, D. W. & Bugg, C. E. (1980) *J. Am. Chem. Soc.* **102**, 5502–5505.
20. Urry, D. W., Okamoto, K., Harris, R. D., Hendrix, C. F. & Long, M. M. (1976) *Biochemistry* **15**, 4083–4089.
21. Urry, D. W. (1983) *Ultrastruct. Pathol.* **4**, 227–251.
22. Urry, D. W. & Long, M. M. (1977) in *Elastin and Elastic Tissue*, eds. Sandberg, L. B., Gray, W. R., & Franzblau, C. (Plenum Press, New York) *Adv. Exp. Med. Biol.* **79**, 685–714.
23. Volpin, D., Urry, D. W., Pasquali-Ronchetti, I., & Gotte, L. (1976) *Micron* **7**, 193–198.
24. Sciortino, F., Prasad, K. U., Urry, D. W. & Palma, M. U. (1988) *Chem. Phys. Ltrs.* **153**, 557–559.
25. Sciortino, F., Palma, M. U., Urry, D. W. & Prasad, K. U. (1988) *Biochem. Biophys. Res. Commun.* **157**, 1061–1066.
26. Sciortino, F., Urry, D. W., Palma, M. U. & Prasad, K. U. (1990) John Wiley & Sons, Inc., *Biopolymers* **29**, 1401–1407.
27. Urry, D. W. (1974) in *Arterial Mesenchyme and Arteriosclerosis*, eds. Wagner, W. D. & Clarkson, T. B. (Plenum Publishing Corp., New York) *Adv. Exp. Med. Biol.* **43**, 211–243.
28. Urry, D. W., Cunningham, W. D. & Ohnishi, T. (1974) *Biochemistry* **13**, 609–616.
29. Urry, D. W., Mitchell, L. W. & Ohnishi (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3265–3269.
30. Khaled, M. A. & Urry, D. W. (1976) *Biochem. Biophys. Res. Commun.* **70**, 485–491.
31. Urry, D. W., Trapane, T. L., Sugano, H. & Prasad, K. U. (1981) *J. Am. Chem. Soc.* **103**, 2080–2089.
32. Venkatachalam, C. M., Khaled, M. A., Sugano, H. & Urry, D. W. (1981) *J. Am. Chem. Soc.* **103**, 2372–2379.
33. Venkatachalam, C. M. & Urry, D. W. (1981) *Macromolecules* **14**, 1225–1229.
34. Urry, D. W., Chang, D. K., Krishna, R., Huang, D. H., Trapane, T. L. & Prasad, K. U. (1989) *Biopolymers* **28**, 819–833.
35. Chang, D. K., Venkatachalam, Prasad, K. U. & Urry, D. W. (1989) *J. of Biomolecular Structure & Dynamics* **6**, 851–858.
36. Wasserman, Z. R. & Salemme, F. R. (1990) *Biopolymers* **29**, 1613–1631.
37. Urry, D. W., Venkatachalam, C. M., Long, M. M. & Prasad, K. U. (1982) in *Conformation in Biology*, eds. Srinivasan, R. & Sarma, R. H. (Ramachandran, G. N. Festschrift Volume, Adenine Press, USA), pp. 11–27.
38. Urry, D. W., Trapane, T. L., Iqbal, M., Vankatachalam, C. M. & Prasad, K. U. (1985) *Biochemistry* **24**, 5182–5189.

39. Urry, D. W., Trapane, T. L., McMichens, R. B., Iqbal, M., Harris, R. D. & Prasad, K. U. (1986) *Biopolymers* **25**, S209–S228.
40. Urry, D. W., Haynes, B. & Harris, R. D. (1986) *Biochem. Biophys. Res. Commun.* **141**, 749–755.
41. Luan, C-H., Jaggard, J., Harris, R. D. & Urry, D. W. (1989) *Int. J. of Quantum Chem., Quantum Biol. Symp.* **16**, 235–244.
42. Edsall, J. T. & McKenzie, H. A. (1983) *Adv. Biophys.* **16**, 53–183.
43. Ben-Naim, A. (1980) in *Hydrophobic Interactions* (Plenum Press, New York).
44. Tanford, C. (1980) in *The Hydrophobic Effect: Formation of Micelles and Biological Membranes* (John Wiley & Sons, New York).
45. Kauzman, W. (1959) *Adv. Protein Chem.* **14**, 1–63.
46. Frank, H. S. & Evans, M. W. (1945) *J. Chem. Phys.* **13**, 507–532.
47. Edsall, J. T. (1935) *J. Am. Chem. Soc.* **57**, 1506–1507.
48. Urry, D. W., Prasad, K. U., Trapane, T. L., Iqbal, M., Harris, R. D. & Henze, R. (1985) *Polymeric Materials: Sci. and Engineering* **53**, 241–245.
49. Thomas, Jr., G. J., Prescott, B. & Urry, D. W. (1987) *Biopolymers* **26**, 921–934.
50. Urry, D. W., Shaw, R. G. & Prasad, K. U. (1985) *Biochem. Biophys. Res. Commun.* **130**, 50–57.
51. Lyerla, J. R. & Torchia, D. A. (1975) *Biochemistry* **14**, 5175–5183.
52. Henze, R. & Urry, D. W. (1985) *J. Am. Chem. Soc.* **107**, 2991–2993.
53. Venkatachalam, C. M. & Urry, D. W. (1986) *Int. J. Quantum Chem., Quantum Biol. Symp.* **12**, 15–24.
54. Buchet, R., Luan, C-H, Prasad, K. U., Harris, R. D. & Urry, D. W. (1988) *J. Phys. Chem.* **92**, 511–517.
55. Urry, D. W. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1610–1614.
56. Go, N. & Okuyama, K. (1976) *Macromolecules* **9**, 867–868.
57. Go, N. & Scheraga, H. A. (1970) *Macromolecules* **3**, 178–187.
58. Go, N. & Scheraga, H. A. (1973) *Macromolecules* **6**, 273–281.
59. Flory, P. J., Ciferri, A., & Hovee, C. A. J. (1960) *J. Polym. Sci.* **XLV**, 235–236.
60. Dorrington, K. L. & McCrum, N. G. (1977) *Biopolymers* **16**, 1201–1222.
61. Urry, D. W., Trapane, T. L. & Prasad, K. U. (1985) *Biopolymers* **24**, 2345–2356.
62. Flory, P. J. (1968) *Rubber Chem. Technol.* **41**, G41–G48.
63. Mark, J. E. (1973) *Rubber Chem. Technol.* **46**, 593–618.
64. Mark, J. E. (1976) *J. Polym. Sci., Macromol. Rev.* **11**, 135–159.
65. Mandelkern, L. (1983) *An Introduction to Macromolecules*, 2nd ed., (Springer-Verlag, New York).
66. Flory, P. J. (1953) *Principles of Polymer Chemistry*, (Cornell University Press, Ithaca, New York).
67. Weis-Fogh, T. & Andersen, S. O. (1970) *Nature (London)* **227**, 718–721.
68. Gosline, J. M. (1978) *Biopolymers* **17**, 677–695.
69. Gosline, J. M. (1980) in *The Mechanical Properties of Biological Materials*, eds. Vincent, J. F. V. & Currey, J. D. (Cambridge University Press, London), pp. 331–357.
70. Gray, W. R., Sandberg, L. B. & Foster, J. A. (1973) *Nature (London)* **246**, 461–466.
71. Urry, D. W. & Venkatachalam, C. M. (1983) *Int. J. Quantum Chem., Quantum Biol. Symp.* **10**, 81–93.
72. Urry, D. W., Venkatachalam, C. M., Wood, S. A. & Prasad, K. U. (1985) in *Structure and Motion: Membranes, Nucleic Acids and Proteins*, eds. Clementi, E., Corongiu, G., Sarma, M. H. & Sarma, R. H., (Adenine Press, Guilderland, New York), pp. 185–203.
73. Urry, D. W. (1985) in *Biomolecular Stereodynamics III*, eds. Sarma, R. H. & Sarma, M. H., (Adenine Press, Guilderland, New York), pp. 173–196.
74. Chang, D. K. & Urry, D. W. (1989) *J. Comput. Chem.* **10**, 850–855.
75. Eyring, H. (1932) *Phys. Rev.* **39**, 746–748.
76. DeBolt, L. & Mark, J. E. (1984) *Polym. Prepr.* **25**, 193–194.
77. DeBolt, L. C. & Mark, J. E. (1987) *Polymer* **28**, 416–422.
78. Andrady, A. L. & Mark, J. E. (1980) *Biopolymers* **19**, 849–855.
79. Rapaka, R. S., Okamoto, K. & Urry, D. W. (1978) *Int. J. Pept. Protein Res.* **11**, 109–127.
80. Urry, D. W., Harris, R. D., Sugano, H., Long, M. M. & Prasad, K. U. (1987) in *Advances in*

- Biomedical Polymers*, ed. Gebelein, C. G. (Plenum Publishing Company, New York), pp. 335–354.
81. Hoeve, C. A. J. & Flory, P. J. (1958) *J. Am. Chem. Soc.* **80**, 6523–6526.
 82. Hoeve, C. A. J. & Flory, P. J. (1974) *Biopolymers* **13**, 677–686.
 83. Urry, D. W., Trapane, T. L., Long, M. M. & Prasad, K. U. (1983) *J. Chem. Soc., Faraday Trans.* **179**, 853–868.
 84. Urry, D. W., Long, M. M., Harris, R. D. & Prasad, K. U. (1986) *Biopolymers* **25**, 1939–1953.
 85. Urry, D. W., Chang, D. K., Zhang, H. & Prasad, K. U. (1988) *Biochem. Biophys. Res. Commun.* **153**, 832–839.
 86. Urry, D. W., Trapane, T. L., Wood, S. A., Walker, J. T., Harris, R. D. & Prasad, K. U. (1983) *Int. J. Pept. Protein Res.* **22**, 164–175.
 87. Onsager, L. (1936) *J. Am. Chem. Soc.* **58**, 1486–1493.
 88. Brooks, B. R., Bruccoleri, R. E., Olafson, B. O., States, D. J., Swaminathan, S. & Karplus, M. (1984) *J. Comput. Chem.* **4**, 187–217.
 89. Karplus, M. & McCammon, J. A. (1983) *Annu. Rev. Biochem.* **53**, 263–300.
 90. Karplus, M. & Kushick, J. N. (1981) *Macromolecules* **14**, 325–332.
 91. Weiner, P. K. & Kollman, P. A. (1981) *J. Comput. Chem.* **2**, 287–303.

Use of Synthetic Peptides in the Structural and Functional Studies of Apolipoprotein A-I

Y. V. Venkatachalapathi, Jan G. Dohlman[†], Warren D. Blackburn^{†*}, Kiran B. Gupta, Shamala K. Srinivas, Jere P. Segrest, and G. M. Anantharamaiah

Department of Medicine, The Atherosclerosis Research Unit and The Division of Rheumatology[†]. UAB Medical Center and V. A. Hospital*. University of Alabama. Birmingham, AL 35294, USA.

The basic rules for predicting the conformations of biologically active peptides and proteins which were initiated by G. N. Ramachandran and his colleagues twenty-five years ago, marked the beginning of an era for the protein structure–function prediction. Our laboratory has been involved in theoretical and experimental study of the amphipathic helix, focusing primarily on apolipoprotein apo A-I.

Coronary artery disease is inversely correlated with levels of plasma HDL and the major HDL protein component, apo A-I (1–4). Apo A-I serves in general as a protein detergent to emulsify the lipid components to minimal diameter particles, and is the major cofactor for activation of the plasma enzyme lecithin:cholesterol acyl transferase (LCAT) which converts free cholesterol into cholesteryl ester (4). Other physiological roles for apo A-I in the acute phase inflammatory response have been suggested by the accompanying altered HDL metabolism (5–7).

The entire 198 residue carboxy-terminal domain of apo A-I, representing the fourth exon in the apo A-I gene, seems to have evolved by eight tandem duplications of a primordial sequence of 22 amino acid residues (8). The primordial 22mer is itself the result of the duplication of an even more primordial sequence of 11 amino acid residues (9). In every case but one, the eight tandem 22mer repeats are punctuated either by single proline residues at the amino-terminus of each 22mer or by two unpaired 11mers inserted between the 22mers (figure 1b).

Each tandem 22mer repeat in apo A-I is an amphipathic helix, a secondary structural motif believed to be important for lipid association (10). The 11mer/22mer evolutionary pathway for apo A-I can be explained as a result of the 3.6 amino acid residues per turn structure of an α -helix; 11 residues equals three complete turns of an α helix and, consequently, duplication of an 11 residue amphipathic α helix produces a 22 residue amphipathic helix in which there is little twist between the polar and

For correspondence:

Dr. G. M. Anantharamaiah, Department of Medicine, Atherosclerosis Research Unit, UAB Medical Center, Birmingham, AL 35294, USA, (205) 934–1884.

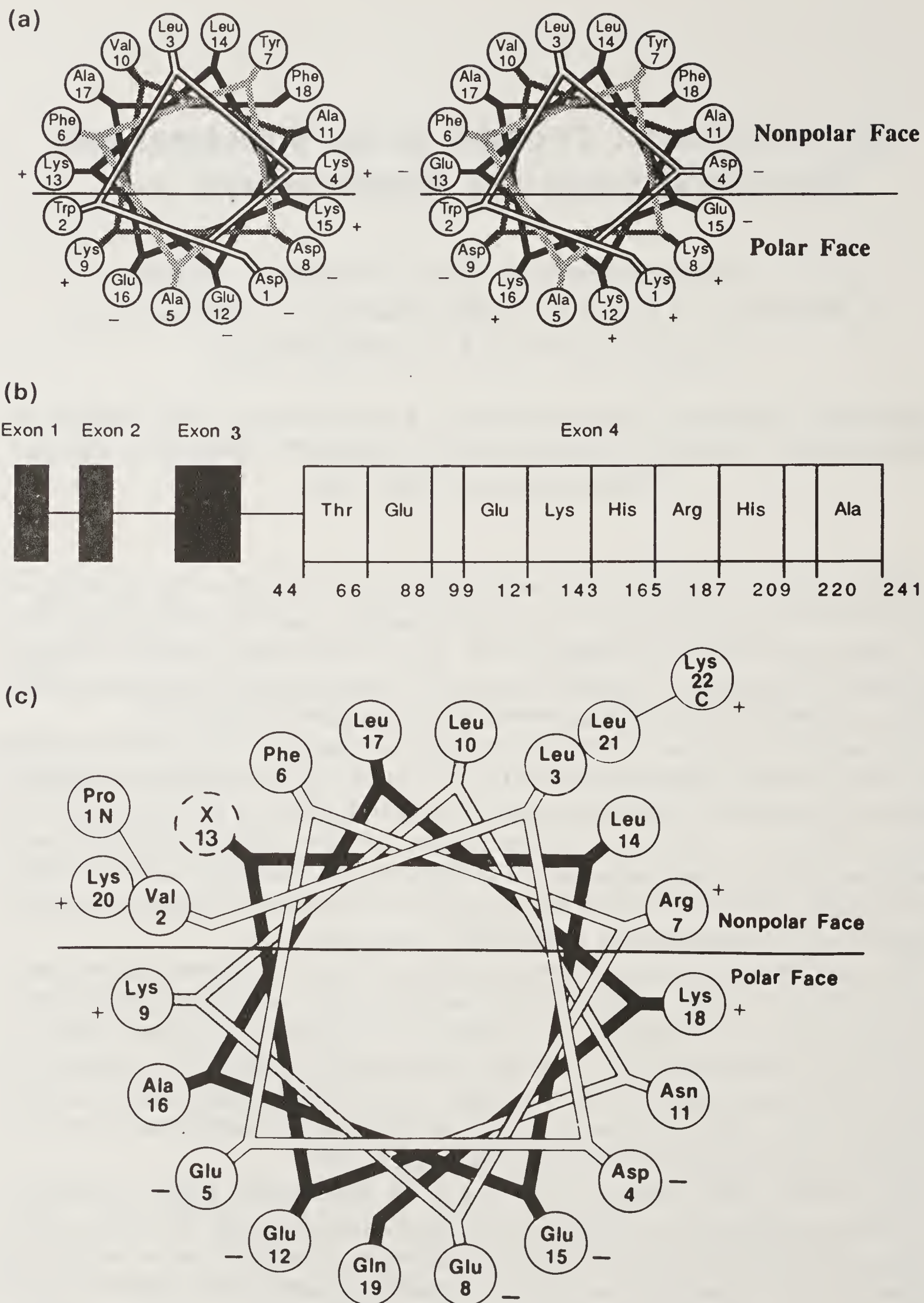


Figure 1. (a) Helical wheel representation of synthetic peptide analogs. 1. Sequence of peptide 18A: Asp Trp Leu Lys Ala. Phe Tyr Asp Lys Val. Ala Glu Lys Leu Lys. Glu Ala Phe. ($L^218A = [Glu^{1,8}, Leu^{11,17}]18A$, $L^318A = [Glu^{1,8}, Leu^{5,11,17}]18A$), Peptide 18R: Lys Trp Leu Asp Ala. Phe Tyr Lys Asp Val. Ala Lys Glu Leu Glu. Lys Ala Phe. ($L^218R = [Glu^{4,9}, Leu^{11,17}]18R$, $L^318R = [Glu^{4,9}, Leu^{5,11,17}]18R$). (b) Genomic structure of Apo A-I. The fourth exon of apo A-I encodes eight 22 aminoacid and two 11 aminoacid amphipathic helical regions. The numbers at

nonpolar faces of the two 11mer halves along the helical axis. The characteristic feature of the amphipathic helical domains of apo A-I as opposed to other classes is the presence of positively charged residues at the polar–nonpolar interface and negatively charged residues at the center of the polar face (11).

As part of a study to understand the molecular properties of apo A-I, we have studied the lipid associating and LCAT activating properties of two sets of synthetic analogs of the amphipathic helix. The first set was apolipoprotein-mimicking peptide analogs which were designed with positively charged residues at the polar–nonpolar interface and negatively charged residues at the center of the polar face. With a view to understand the importance of such a charge distribution, we compared the properties of these peptides with peptides of the opposite charge distribution. That is, negatively charged residues at the polar–nonpolar interface and positively charged residues at the center of the polar face (figure 1). In order to understand the role of proline in repeating 22mer of apolipoproteins, a dimer (18A–Pro–18A, 37pA) was also synthesized.

With a view to mimic closely the tandem 22mer in apo A-I, a consensus sequence approximation of the primordial 22mer, called A-I_{con}, was derived from the amino acid sequence of human apo A-I: Pro Val Leu Asp Glu·Phe Arg Glu Lys Leu·Asn Glu X Leu Glu·Ala Leu Lys Gln Lys·Leu Lys. This sequence was constructed by selection of the most prevalent residue at each position. Compared to the other 21 residues in the consensus sequence, residue 13, denoted by X, has a marked and unusual amino acid polymorphism. When the consensus sequence is organized as a helical wheel (figure 1c), all amino acid residues are radially distributed as a typical apolipoprotein amphipathic helix except for residue 13. Although residue 13 is positioned on the nonpolar face of the consensus sequence, six out of the eight tandem repeats have a charged residue at this position and a seventh is polar.

The anomalous nature of residue 13 in the consensus sequence (figure 1c) suggested to us the possibility that residue 13 in several or all of the tandem 22mer repeats might play a role in the LCAT-activating properties of apo A-I. In order to investigate this possibility, four 22mer peptide analogs of the consensus sequence differing only in the thirteenth residue, and four 44mer peptides representing three homo- and one heterodimeric combinations of the four 22mer peptides, were synthesized. We describe here the properties of these peptides with respect to: (a) their relative affinity for phospholipid, (b) their relative ability to activate the enzyme LCAT using two different egg lecithin assay systems, and (c) their effect on immunoglobulin-mediated neutrophil activation.

Materials and Methods

Apo A-I was purified as described earlier (12). The enzyme LCAT was purified according to the method described (13).

the bottom indicate the position of aminoacids and the aminoacid residues in the boxes in fourth exon are the 13th residues present in the 22mers of the tandem repeats. (c) Helical wheel representation of consensus peptide for 22mer tandem repeats of A-I: Pro Val Leu Asp Glu. Phe Arg Glu Lys Leu. Asn Glu X Leu Glu. Ala Leu Lys Gln Lys. Leu Lys.

Peptide Synthesis

Peptides were synthesized by the solid phase synthesis procedure as described elsewhere (14). The crude peptides were dissolved in 6 M guanidine hydrochloride and dialysed against water (Spectrapor dialysis membrane, 1000 MW cut off), lyophilized and purified using HPLC. The sequence and purity of peptides were confirmed by amino acid analysis, peptide sequencing and analytical HPLC. Peptides were dissolved in guanidine hydrochloride (6 M) to obtain a concentration of about 5 mg/ml. These solutions were dialysed against phosphate buffer (pH 7.4) using 1000 MW cutoff bags to ensure the proper pH of solutions of peptides. The concentration of solutions were determined by quantitative amino acid analysis.

Electron Microscopy of Peptide:DMPC Complexes

Peptides complexed with multilammellar vesicles of dimyristoyl phosphatidylcholine (DMPC) (Avanti Polar Lipids, Inc., Birmingham, AL) at a 1:2.5 (peptide:DMPC) weight ratio were stained with 2% potassium phosphotungstate, pH 5.9, and examined with a Philips EM400 microscope on carbon-coated Formvar grids.

LCAT Activation

Egg PC vesicular assay system: Egg PC vesicles containing unesterified cholesterol prepared earlier were used to determine the percent conversion of free cholesterol to cholesteryl ester. The assay was conducted by incubating a mixture of egg lecithin (45 nM) and cholesterol (10 nM), EDTA (1 mM), 2-mercaptoethanol (4 mM), bovine serum albumin (22 μ M), (pH 7.4) and varying amounts of activator (peptide or apo A-I) in a total volume of 250 μ l for 1 h at 37°C. The reaction was initiated by adding 50 μ l of suitably diluted purified enzyme and incubating for 1 h at 37°C. Free cholesterol and cholesteryl ester were separated by thin layer chromatography on silica gel. The percent conversion of cholesterol to ester was calculated for increasing amounts of activator.

Preparation of peptide or protein:egg PC discoidal particles: Egg PC containing [3 H] cholesterol (9:2 molar ratio) was vortexed and mixed with apo A-I or peptides at a peptide:egg PC ratio of 1:2 (W:W), the mixture solubilized in an excess of 3% sodium cholate (approx. 300 μ l), and dialyzed against TRIS buffer overnight using 50,000 MW cut off. Analysis of the concentration of proteins (by amino acid analysis), cholesterol (by radiolabel remaining), and phospholipid (15) indicated no loss of any of them.

Egg PC discoidal assay system: The peptide/protein:egg PC complexes were incubated with purified LCAT at 37°C for varying periods of time. Cholesteryl ester formation was assayed as described for the sonicated vesicles.

Neutrophil assays: Purified human neutrophils were isolated by dextran sedimentation and density centrifugation as described (16). Fc-receptor mediated

neutrophil activation was measured in the presence of apo A-I or amphipathic peptides by the levels of superoxide production and degranulation (17, 18).

Results

As one means of determining relative affinity for phospholipid, the different consensus peptides were mixed with multilamellar vesicles of DMPC and their ability to convert the multilamellar vesicles to minimal size discoidal particles was determined. The conversion to discoidal particles was followed by light scattering analysis, negative stain electron microscopy, and nondenaturing gradient gel electrophoresis. The results of the electron microscopic studies of the peptide-DMPC complexes for the different peptides are consistent with these and were found by light scattering and gradient gel analyses (data not shown). As determined in earlier studies from this laboratory (19), the diameter of the discoidal peptide-DMPC particles vary inversely with the lipid affinity of the peptide analogs.

Although the consensus sequence analogs differ by only one amino acid residue in the thirteenth position, they have widely varying lipid affinities as judged by the

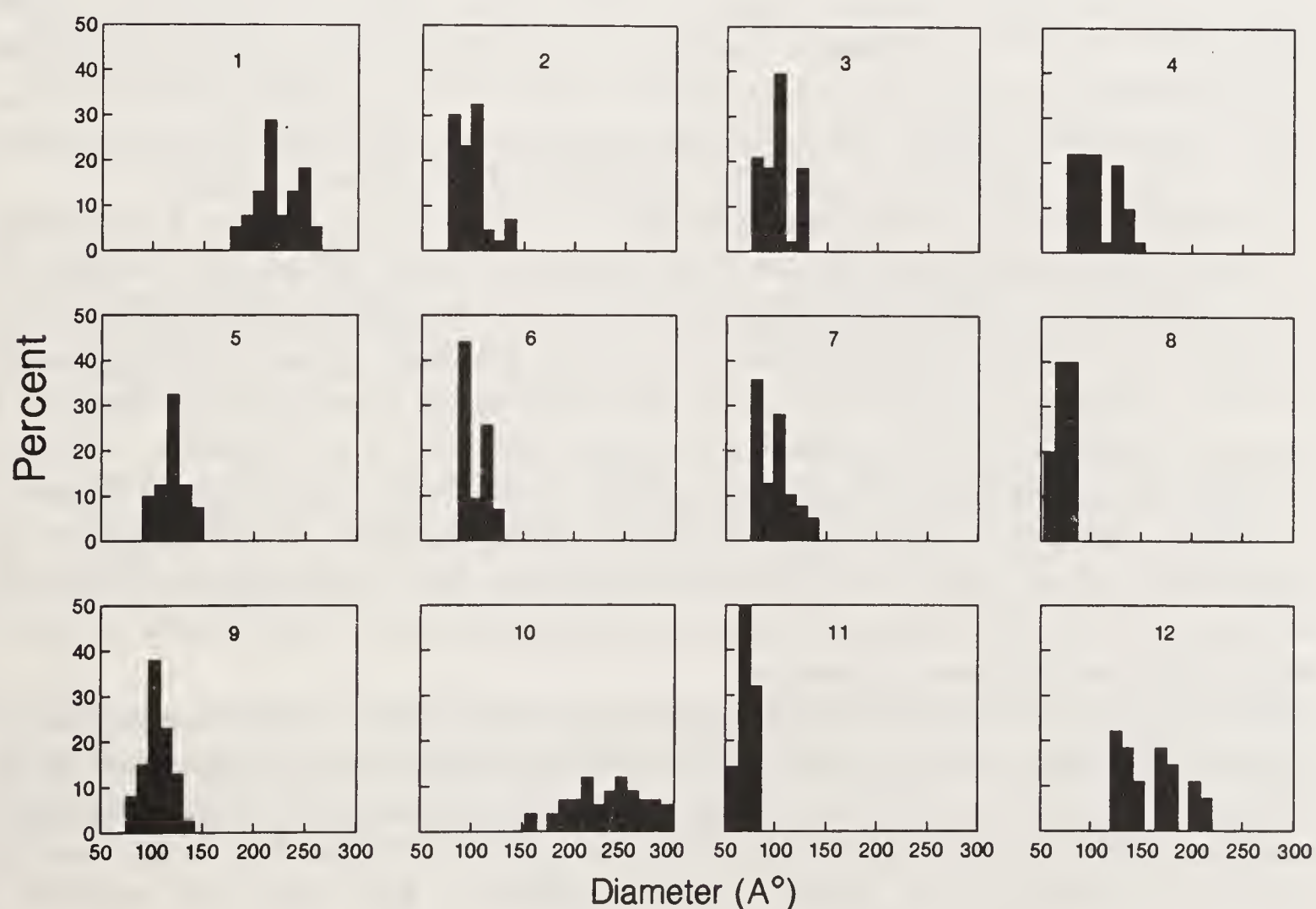


Figure 2. Morphographic analysis of diameters of discoidal complexes of analog peptides with DMPC. 1. [Glu¹³]A-I_{con}, 2. [Ala¹³]A-I_{con}, 3. [Arg¹³]A-I_{con}, 4. [His¹³]A-I_{con}, 5. ([Glu¹³]A-I_{con})₂, 6. [Arg¹³]A-I_{con}-[Glu¹³]A-I_{con}, 7. ([Arg¹³]A-I_{con})₂, 8. ([His¹³]A-I_{con})₂, 9. 18A, 10. [Glu^{1,8},Leu^{11,17}]18A, 11. 18R and 12. [Glu^{4,9},Leu^{11,17}]18R

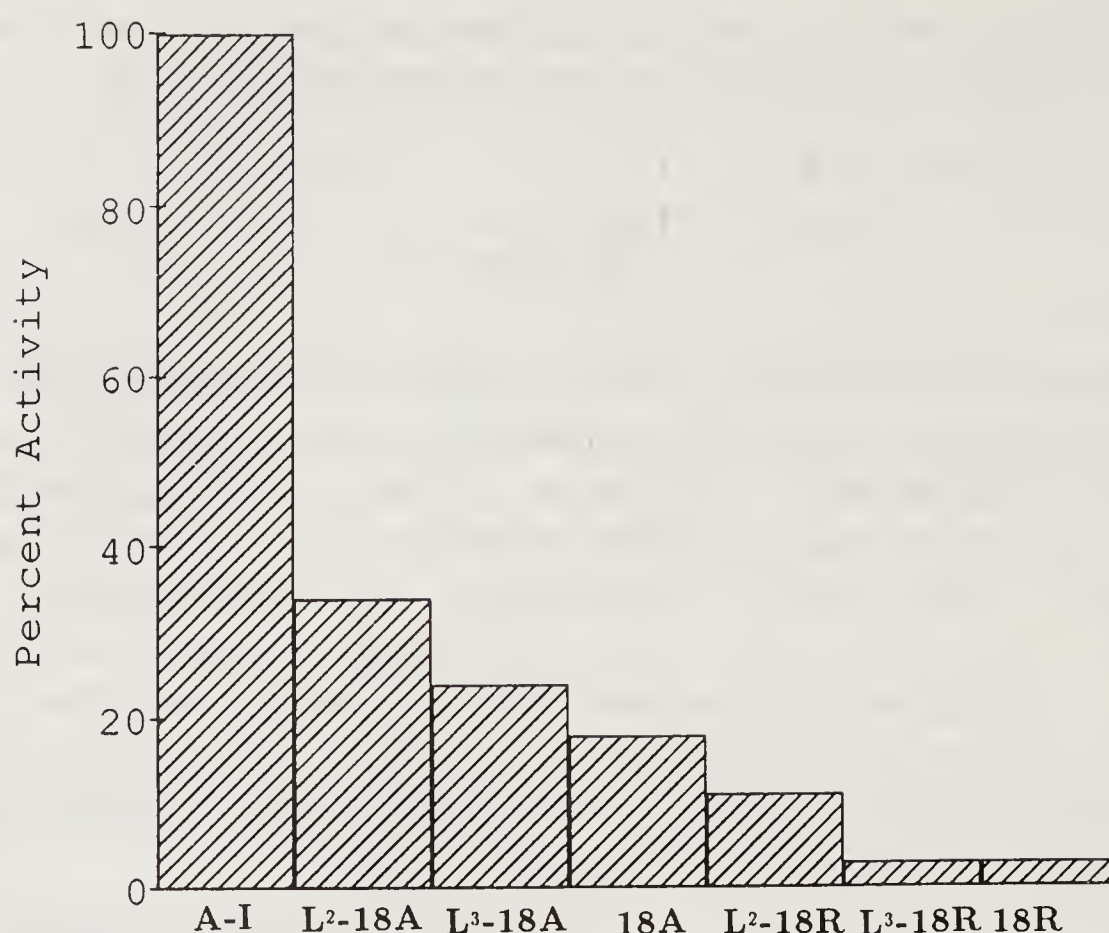


Figure 3. Comparison of LCAT activating ability of 18A and 18R analogs with apo A-I.

size of the discoidal complexes (figure 2). The 44mer dimer analogs had a higher affinity than the 22mer monomer analogs. Approximate rank order lipid affinities for the dimers were $([\text{His}^{13}] \text{A-I}_{\text{con}})_2$ (HH) $> ([\text{Arg}^{13}] \text{A-I}_{\text{con}})_2$ (RR) $> [\text{Arg}^{13}] \text{A-I}_{\text{con}} - [\text{Glu}^{13}] \text{A-I}_{\text{con}}$ (RE) $> ([\text{Glu}^{13}] \text{A-I}_{\text{con}})_2$ (EE), and the rank order among monomers was $[\text{Ala}^{13}] \text{A-I}_{\text{con}}$ (A) $> [\text{His}^{13}] \text{A-I}_{\text{con}}$ (H) $> [\text{Arg}^{13}] \text{A-I}_{\text{con}}$ (R) $> [\text{Glu}^{13}] \text{A-I}_{\text{con}}$ (E).

The results of the LCAT activation studies are shown in figures 3 and 4. In general, the 18 A analogs were more effective LCAT activators than 18R analogs. Among the A-I_{con} peptides, the 22mer monomers were found to be poorer activators of LCAT than the 44mer dimers. The R monomer and the RR dimer produce approximately the same degree of esterification. Both are considerably more effective than the E monomer; however, the EE dimer is the most effective of the peptides. At lower activator concentrations, EE equalled apo A-I activation at equilibrium (figure 4) and RE equalled apo A-I at higher activator concentrations. Examination of the incubation mixtures of all of the consensus peptides with egg lecithin SUV showed that none of the eight analogs formed discoidal particles from egg lecithin at any of the peptide concentrations examined.

In order to test the 44mer peptides in the discoidal system, discoidal complexes of peptide:egg lecithin were prepared by the cholate dialysis method described in the Methods section, and were used as substrates for the enzyme LCAT (20). Under these conditions, EE has an esterification rate comparable to apo A-I, both reaching apparent equilibrium in approx. 20 min, whereas RE had not reached a comparable level of esterification by 45 min (figure 4). HH produces an apparent equilibrium esterification of only 40% that of apo A-I, consistent with a nonspecific amphipathic helix effect.

The ability of apo A-I to modulate neutrophil Fc-receptor mediated activation was measured via superoxide production and degranulation. Apo A-I inhibited both

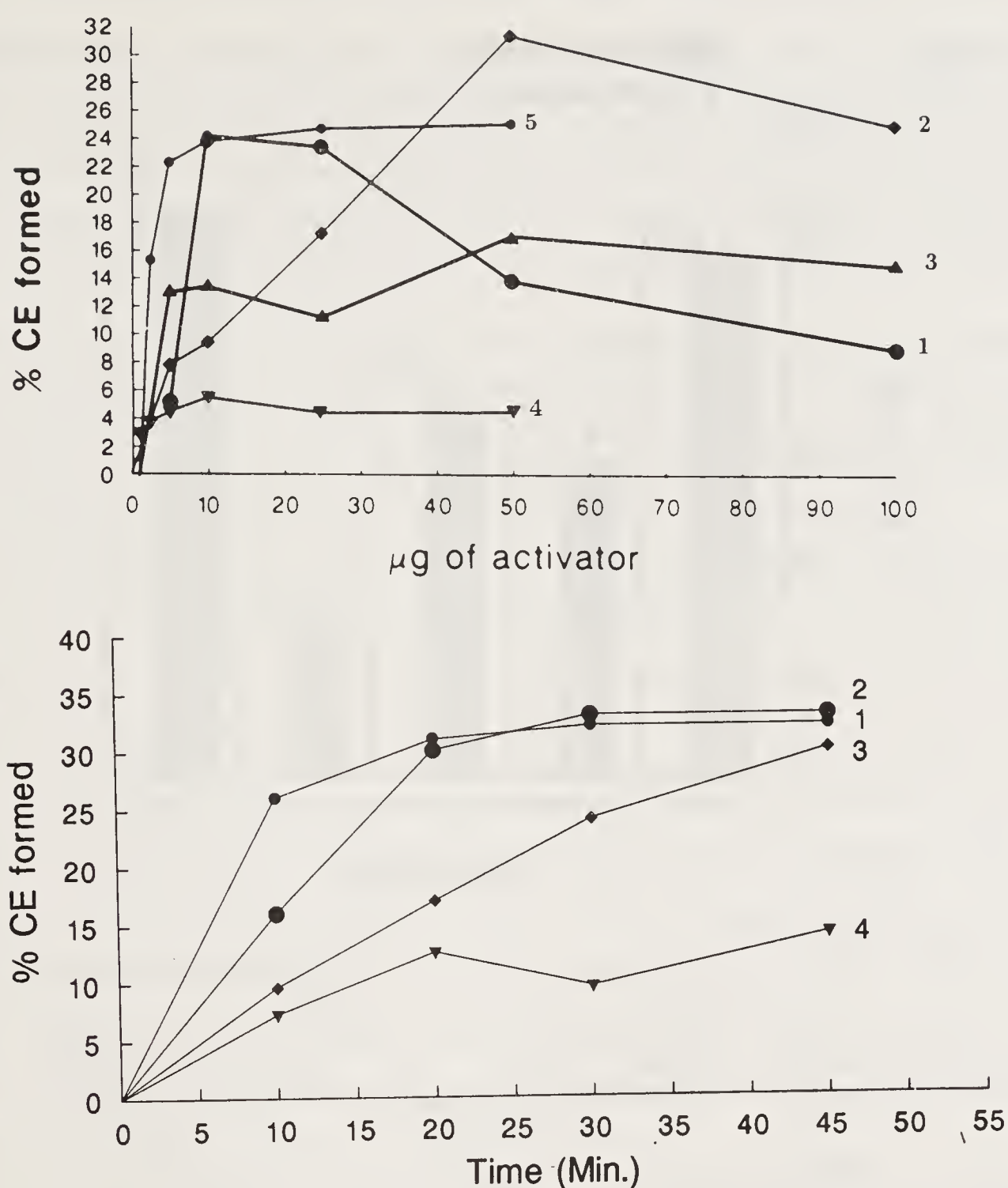


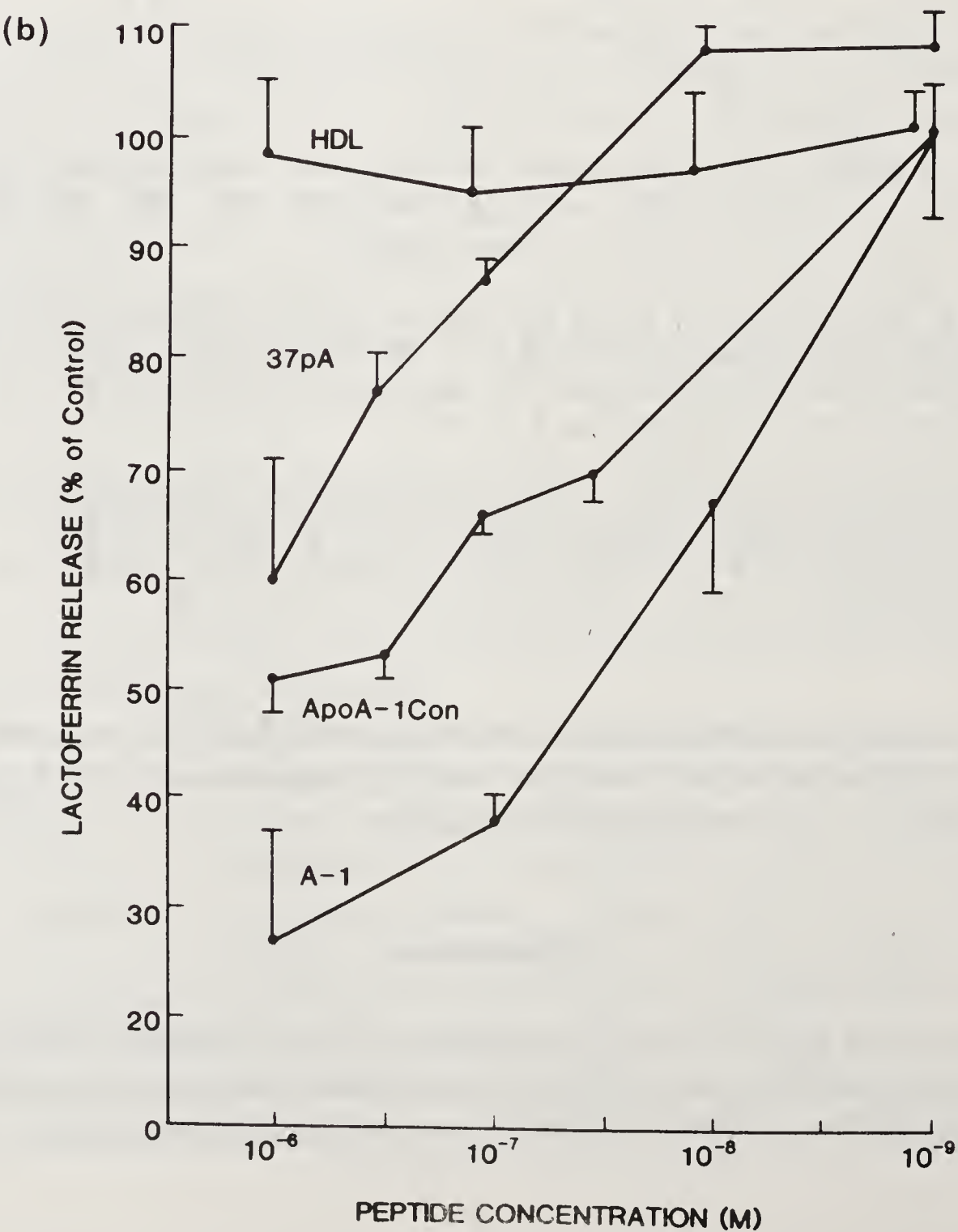
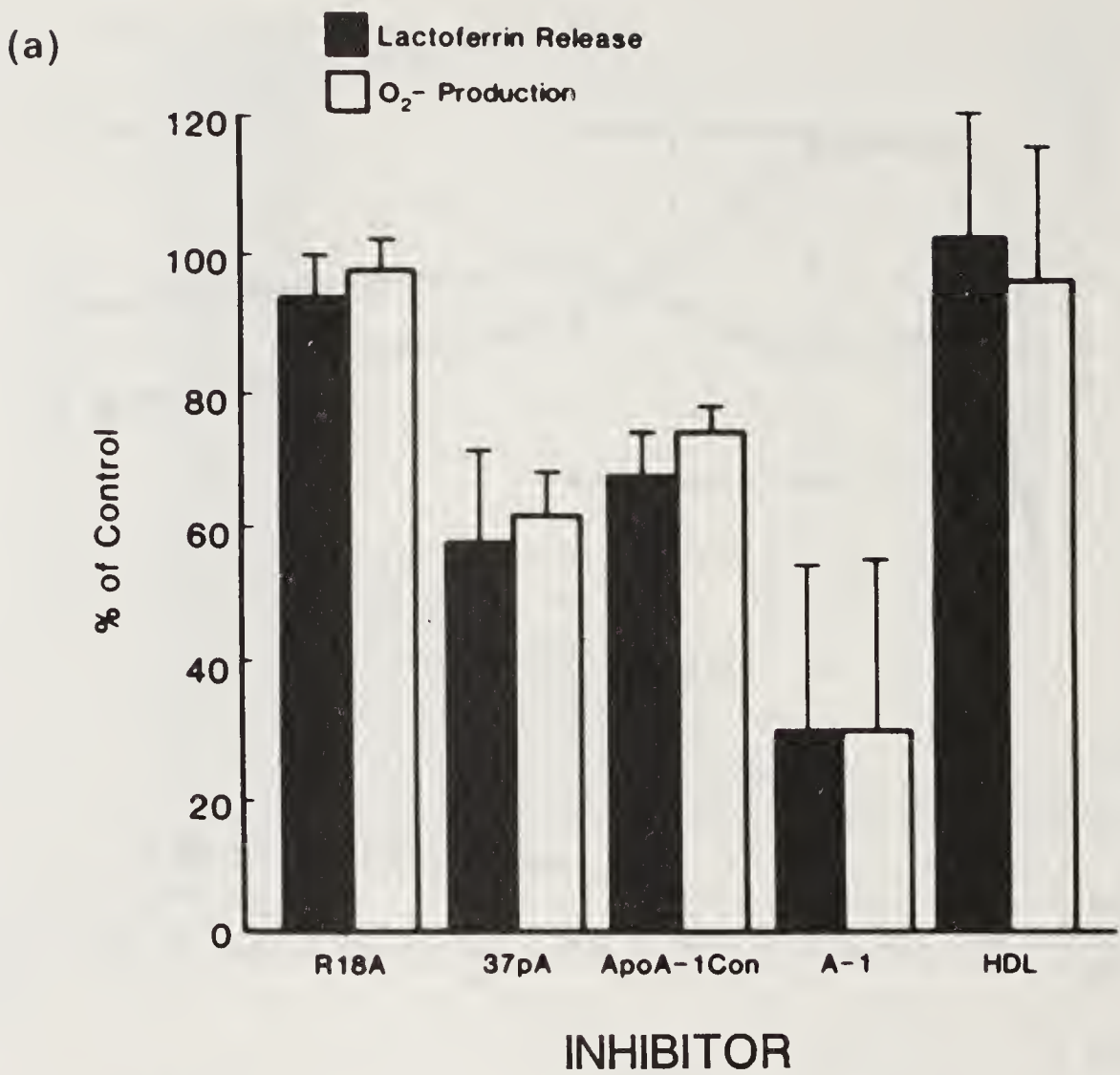
Figure 4. Top Activation of LCAT by synthetic peptide analogs as measured by the egg lecithin SUV procedure (20). 1. (●—●), $([\text{Glu}^{13}]\text{A-I}_{\text{con}})_2$; 2. (◆—◆), $[\text{Arg}^{13}]\text{A-I}_{\text{con}}-[\text{Glu}^{13}]\text{A-I}_{\text{con}}$; 3. (▲—▲), $([\text{Arg}^{13}]\text{A-I}_{\text{con}})_2$; 4. (▼—▼), $([\text{His}^{13}]\text{A-I}_{\text{con}})_2$; 5. (●—●), Apo A-I.

Bottom Activation of LCAT by synthetic peptide analogs as measured by the egg lecithin discoidal particle procedure (20). 1. (●—●), apo A-I; 2. (●—●), $([\text{Glu}^{13}]\text{A-I}_{\text{con}})_2$; 3. (◆—◆), $[\text{Arg}^{13}]\text{A-I}_{\text{con}}-[\text{Glu}^{13}]\text{A-I}_{\text{con}}$; 4. (▼—▼), $([\text{His}^{13}]\text{A-I}_{\text{con}})_2$.

parameters at well below physiological concentrations (figure 5). Two peptide analogs which most effectively mimic the lipid binding properties of apo A-I were also inhibitory whereas a control peptide was inactive.

Discussion

The theory of the lipid-associating amphipathic helix originated with the proposal by our laboratory in 1974 that members of the class of exchangeable apolipoproteins contain amphipathic α -helical domains involved in lipid-association (13). The



amphipathic helix is defined as an α helix with opposing polar and nonpolar faces. Based on our initial observation, amphipathic helical domains have been described for other putative lipid-associating proteins, such as certain polypeptide hormones (21). The amphipathic helix class associated with the apolipoproteins, as opposed to that associated with polypeptide hormones, is unique. Positively-charged residues cluster at the polar–nonpolar interface and negatively-charged amino acid residues cluster at the center of the polar face.

Our laboratory has been using synthetic peptide analogs for a better understanding of the role played by charged residues in lipid-affinity and LCAT activating properties of the amphipathic helix. The 18A and 18R analogs were synthesized and studied to understand the importance of the presence of positively charged residues at the polar–nonpolar interface of the apolipoprotein-class amphipathic helix.

The histograms in figure 2 show the sizes of peptide:DMPC complexes. The 18A analog formed smaller discoidal complexes compared to the 18R analog. This was found to be true for the other 18A and 18R peptide analogs in which helicity was increased by the substitution of alanines with leucines and aspartic with glutamic residues (22,23). These results indicate that the analogs with positively charged residues at the polar–nonpolar interface have greater lipid affinity. The fact that this is not a function of differences in the helix-dipole moment between 18A and 18R analogs was confirmed by the fact that Ac-18A-NH₂ had increased lipid affinity compared to Ac-18R-NH₂ (unpublished results). These results thus confirm the importance of positively charged residues at the polar–nonpolar interface of apolipoprotein class amphipathic helices for lipid affinity.

The results of LCAT activation are shown in figures 3 and 4. Compared to 18A analogs, 18R analogs have less LCAT activating ability, thus suggesting that the position of charged residues is also important for the ability of an amphipathic peptide analog to serve as a cofactor of the enzyme LCAT.

Since the apolipoprotein class amphipathic peptides possessed 25–30% LCAT activating ability of apo A-I, we designed a closer peptide mimic for the lipid-associating domains of apo A-I. Because each tandem 22mer in apo A-I appears to represent a duplication of a primordial gene sequence, as part of our studies of the molecular properties of amphipathic helices, a consensus sequence approximation of the primordial 22mer, called A-I_{con}, was derived from the amino acid sequence of human apo A-I. A consensus nucleotide sequence has been proposed for apo A-I

Figure 5. The effect of HDL, apo A-I, and model amphipathic peptides on surface bound IgG mediated neutrophil activation. (a) The effect of 10^{-6} M HDL, apo A-I, or peptide ($n = 8$). (b) Dose dependent effect of HDL, apo A-I and model amphipathic peptides on surface bound IgG-mediated neutrophil activation ($n = 6$). Purified human neutrophils from normal donors were incubated for 15 min with either apo A-I or test peptides at 37°C. Cell suspensions were then added to microtiter wells that had been precoated with human IgG. After a 10 min incubation at 37°C, cell free supernates were assessed for superoxide using a colorimetric cytochrome reduction assay as described (18). Degranulation was measured after a 1 h incubation at 37°C by solid phase radioimmunoassay, of the cell free supernates, for lactoferrin (19).

(24). The peptide derived from this consensus sequence does not have an amphipathic helical structure and therefore, the consensus amphipathic helical sequence was derived from the 22mer repetitive amino acid sequence.

CNBr fragments of apo A-I have been subjected to LCAT activation studies (25). There are three Met residues in the human apo A-I sequence at positions 86, 112 and 148. CNBr degradation, thus results in two larger and two smaller fragments. The two larger CNBr fragments of apo A-I have been shown to activate LCAT only to the extent of 20% to 25% of apo A-I. The smaller two CNBr fragments have been shown to have essentially no LCAT activating ability. This approach has failed to localize LCAT activating domain(s) of apo A-I. It is therefore possible that this domain could be a part of two Met containing sequence of apo A-I.

The enzyme LCAT cleaves the acyl chain at the Sn-2-position of a glycerophospholipid molecule and then transfers the chain to the hydroxyl group of cholesterol. The products, free fatty acid and cholesteryl ester, therefore depend on the environment of the acceptor hydroxyl group. It is possible that not all the free fatty acid product is converted to cholesteryl ester. For the sake of the simplicity of studying apo A-I mimicry, we choose to follow the conversion of labelled free cholesterol to its cholesteryl ester rather than following the formation of free fatty acid.

Our present results indicate that, in two different assay systems, the EE dimer closely mimics apo A-I as an activator of LCAT. When appropriate criteria are used for comparison, this peptide is the most active LCAT activating peptide analog, synthetic or native fragment, yet described. While the RE dimer can produce cholesteryl ester in comparable amounts to that of EE this occurs only at higher concentrations or after longer incubation periods. The RR dimer has significantly less effect on LCAT than either EE or RE. We therefore speculate that the Glu half of the Arg-Glu dimer is the active portion of this peptide, producing an equal esterification of cholesterol only at higher concentrations or after longer incubation times. Since the 44mer analog EE equals the equilibrium LCAT activating ability of apo A-I at the same concentrations at which apo A-I also reaches its maximum, we conclude that the EE dimer is a close mimic of the LCAT activating ability of apo A-I in the SUV system. As noted, the only other peptide analog that equals the LCAT activating ability of apo A-I in the SUV system, RE does so only at higher concentrations.

The distribution of the 22mer tandem repeats in the carboxy terminal 198 residues of apo A-I (exon 4) is diagrammed in figure 1b. Each repeat is identified by the amino acid residue at the thirteenth position. Repeats two and three (residues 66–121), separated by an unpaired 11mer, contain the only two Glu residues present in the thirteenth position of the eight repeats. As noted earlier, examination of CNBr fragments of apo A-I for LCAT activation have failed to identify a fragment with activity comparable to that of intact A-I (25). It should be noted that two of the three Met residues in apo A-I are located between residues 66 and 121 and that the second Met is immediately adjacent to the second Glu; therefore the putative LCAT activation domain would thus be extensively disrupted by CNBr fragmentation.

We therefore propose that a major LCAT-activating domain of apo A-I is associated with the two, Glu-containing, 22mer tandem repeats located between residues 66 and 121 in the native apolipoprotein. Consistent with this proposal is the conservation of these two Glu in apo A-I from other species (26, 27). Since $([\text{Glu}^{13}] \text{A-I}_{\text{con}})_2$ differs from the native sequence of 66–121 in having an intervening Pro rather than an

unpaired 11mer, and since the analog EE dimer is a consensus sequence rather than the native sequence, studies are presently underway to test the hypothesis further by direct synthesis of 66–121. Recent studies by Jonas *et al.* (28) provide additional support for this hypothesis.

The acute systemic inflammatory response is known to be accompanied by drastically altered HDL metabolism, including an increased production of HDL associated with serum amyloid A (SAA) and a large drop in plasma apo A-I levels (6). It is also known that SAA *in vitro* displaces apo A-I from HDL particles. These data thus suggest a role for HDL and apo A-I in inflammation (5). Neutrophils have a prominent role in acute inflammation and have been shown to interact with HDL (29). Although HDL had no effect on neutrophil activation, we demonstrated profound neutrophil inhibitory effects using well-below physiologic levels of plasma apo A-I. This effect could be mediated by the postulated displaced apo A-I (30) or by the postulated prebeta apo A-I (31). It may also be significant that the acute phase reactant serum amyloid A is known to displace apo A-I from HDL particles *in vitro*. A lack of cell inhibition by lipid-bound apo A-I suggested to us that this effect was also mediated by the amphipathic helical domains. Indeed, the peptide mimics which can effectively bind lipid also were inhibitory.

In conclusion, we propose that the lipid associating domains of apo A-I possess physiological roles such as LCAT activation and neutrophil regulation. Interactions of apo A-I with cell membranes may have general relevance. For example, we have recently found that apo A-I and its mimetic peptide analogs inhibit viral induced cell-fusion, and thus possess antiviral properties (32, 33). These detailed structure–function studies of the amphipathic helixes may also prove to be useful to the understanding of cell-regulatory properties of apo A-I.

Acknowledgement

This work was supported by a grant HL-34343 from the National Institutes of Health.

References

1. Brunzell, J. D., Albers, J. J., Chait, A., Grundy, S. M., Groszek, E. & McDonald, G. B. (1983) *J. Lipid Res.* **24**, 147–155.
2. Maciejko, J. J., Holmes, D. R., Kotte, B. A., Zinsmeister, A. R., Dinh, D. M. & Mao, S. J. T. (1983) *N. Engl. J. Med.* **309**, 385–389.
3. Sniderman, A., Shapiro, S., Marpole, D., Skinner, B., Teng, B. & Kwiterovich, P.O. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 604–608.
4. Kottke, B. A., Zinsmeister, A. R., Holmes Jr, M. D., Kneller, R. W., Hallaway, B. J. & Mao, S. J. (1986) *Apolipoproteins and coronary artery disease. Mayo Clinic Proceedings* **61**, 313–320.
5. Coetzee, G. A., Strachan, A. F., van der Westhuyzen, D. R., Hoppe, H. C., Jenenah, M. S. & de Beer, F. C. (1986) *J. Biol. Chem.* **261**, 9644–9651.
6. Cabana, V. G., Sieggel, J. N. & Sabesin, S. M. (1989) *J. Lipid Res.* **30**, 39–49.
7. Clifton, P. M., Mackinnon, A. M. & Barter, P. J. (1985) *J. Lipid Res.* **26**, 1389–1398.
8. Luo, C. C., Li, W. H., Moore, M. N. & Chan, L. (1986) *J. Mol. Biol.* **187**, 325–340.
9. Fitch, W. M. (1977) *Genetics* **86**, 623–644.

10. Segrest, J. P., Jackson, R. L., Morrisett, J. D. & Gotto Jr, A. M. (1973) *FEBS Lett.* **38**, 247–253.
11. Segrest, J. P., De Loof, H., Dohlman, J. G., Brouillette, C. G. & Anantharamaiah, G. M. (1990) *Proteins* (in press).
12. Hughes, T. A., Moore, M. A., Neame, P., Medley, M. & Chung, B. H (1988) *J. Lipid Res.* **29**, 363–376.
13. Chen, C. H. & Albers, J. J. (1981) *Biochem. Med.* **28**, 215–226.
14. Anantharamaiah, G. M. (1985) *Methods Enzymol.* **128**, 628–668.
15. Yokoyama, S., Fukushima, D., Kupferberg, J. P., Kezdy, F. J, & Kaiser, E. T. (1980) *J. Biol. Chem.* **255**, 7333–7339.
16. Blackburn, W. D., Heck, L. H., Koopman, W. J. & Gresham, H. D. (1987) *Arthritis Rheum.* **30**, 106–114.
17. Cohen, H. J. & Chovanciec, M. E. (1978) *J. Clin. Invest.* **61**, 1081–1087.
18. Blackburn, W. D., Koopman, W. J., Schrohenloher, R. E. & Heck, L. H. (1986) *J. Clin. Invest.* **40**, 347–355.
19. Anantharamaiah, G. M., Jones, J. L., Brouillette, C. G, Schmidt, C. F, Chung, B. H., Hughes, T. A., Bhowan, A. S. & Segrest, J. P. (1985) *J. Biol. Chem.* **260**, 10248–10255.
20. Chung, B. H., Anantharamaiah, G. M., Brouillette, C. G., Nishida, T. & Segrest, J. P. (1985) *J. Biol. Chem.* **260**, 10256–10262.
21. Kaiser, E. T. & Kezdy, F. J. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1137–1143.
22. Epand, R. E., Gawish, A., Iqbal, M., Gupta, K. G., Chen, C. H., Segrest, J. P & Anantharamaiah, G. M. (1987) *J. Biol. Chem.* **262**, 9389–9396.
23. Epand, R. E., Surewicz, W. K., Hughes, D. W., Mantsch, H., Segrest, J. P., Allen, M. T, & Anantharamaiah, G. M. (1989) *J. Biol. Chem.* **264**, 4628–4635.
24. Karathanasis, S. K., Zannis, V. L. & Breslow, J. L. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6147–6151.
25. Fielding, C. J., Shore, V. G. & Fielding, P. E. (1972) *Biochem. Biophys. Res. Commun.* **46**, 1493–1498.
26. Yang, C. Y., Yang, T., Pownall, H. J. & Gotto, A. M. (1986) *Eur. J. Biochem.* **160**, 427–431.
27. Chung, H., Randloph, A., Reardon, I. & Henrikson, R. L. (1982) *J. Biol. Chem.* **257**, 2961–2967.
28. Jonas, A., Kezdy, K. E. & Wald, J. H. (1989) *J. Biol. Chem.* **264**, 4818–4824.
29. Shepard, E. G., De Beer, F. C., De Beer, M. C., Jeenah, M. S., Coetzee, G. A. & van der Westhuyzen, D. R. (1987) *Biochem. J.* **248**, 919–926.
30. Cabana, V. G., Strachan, A. F., Westhuyzen, D. R., Hoppe, H. C., Jeenah, M. S. & de Beer, F. C. (1989) *J. Lipid Res.* **30**, 39–49.
31. Ishida, B. Y., Frolich, J. & Fielding, C. J. (1987) *J. Lipid Res.* **28**, 778–786.
32. Srinivas, R. V., Birkedal, B., Owens, R. J., Anantharamaiah, G. M., Segrest, J. P. & Compans, R. W. (1990) *Virology* **176**, 48–57.
33. Owens, R. J., Anantharamaiah, G. M., Kahlon, J. B., Srinivas, R. V., Compans, R. W. & Segrest, J. P. (1990) *J. Clin. Invest.* (In Press).

Role of Basic Amino Acid Residues in the Amphipathic Helix: the Snorkel Hypothesis

Jere P. Segrest*, Y. V. Venkatachalapathi, Shamala K. Srinivas,
Kiran B. Gupta, Hans De Loof and G. M. Anantharamaiah

Department of Medicine and The Atherosclerosis Research Unit,
UAB Medical Center, Birmingham, AL 35294, USA

The amphipathic helix is an often encountered secondary structural motif in biologically active peptides and proteins. The amphipathic helix was first described as a unique structure–function motif involved in lipid interaction by Segrest *et al.* (1) in 1973. Amphipathic helical domains have also been described for polypeptide hormones, such as endorphins; polypeptide venoms, such as bombolitin; polypeptide antibiotics, such as the magainins; and complex transmembrane proteins, such as bacteriorhodopsin (2). Four of these amphipathic helix classes belong to lipid-associating proteins: apolipoproteins (A-class), certain polypeptide hormones (H-class), lytic polypeptides (L-class) and transmembrane proteins (M-class) (figure 1).

The exchangeable apolipoproteins from plasma lipoproteins (apo A-I, A-II, A-IV, C-I C-II, C-III, and E) have been included in the class A amphipathic helices. This class differs from the other classes in that it possesses positively charged residues at the polar–nonpolar interface and negatively charged residues at the center of the polar face. Our approach to understanding the factors responsible for the lipid association of the apolipoprotein class amphipathic helix was the use of model synthetic peptide analogs.

We designed the following two model peptides: Asp Trp Leu Lys Ala·Phe Tyr Asp Lys Val·Ala Glu Lys Leu Lys·Glu Ala Phe (18A) and Lys Trp Leu Asp Ala·Phe Tyr Lys Asp Val·Ala Lys Glu Leu Glu·Lys Ala Phe (18R) (figure 2a). The peptide 18A was designed to mimic the class A amphipathic helix. In this peptide, positively charged residues were located at the polar–nonpolar interface and negatively charged residues at the center of the polar face. To understand the importance of such an arrangement of charged residues, 18R was designed in which negatively charged Asp and Glu were located at the polar–nonpolar interface and positively charged residues at the center of the polar face. Results from these two peptides showed that the 18A had a higher lipid affinity. This was thought to be due to the presence of positively charged residues at the interface, with longer acyl chain length, that may play an

*For correspondence.

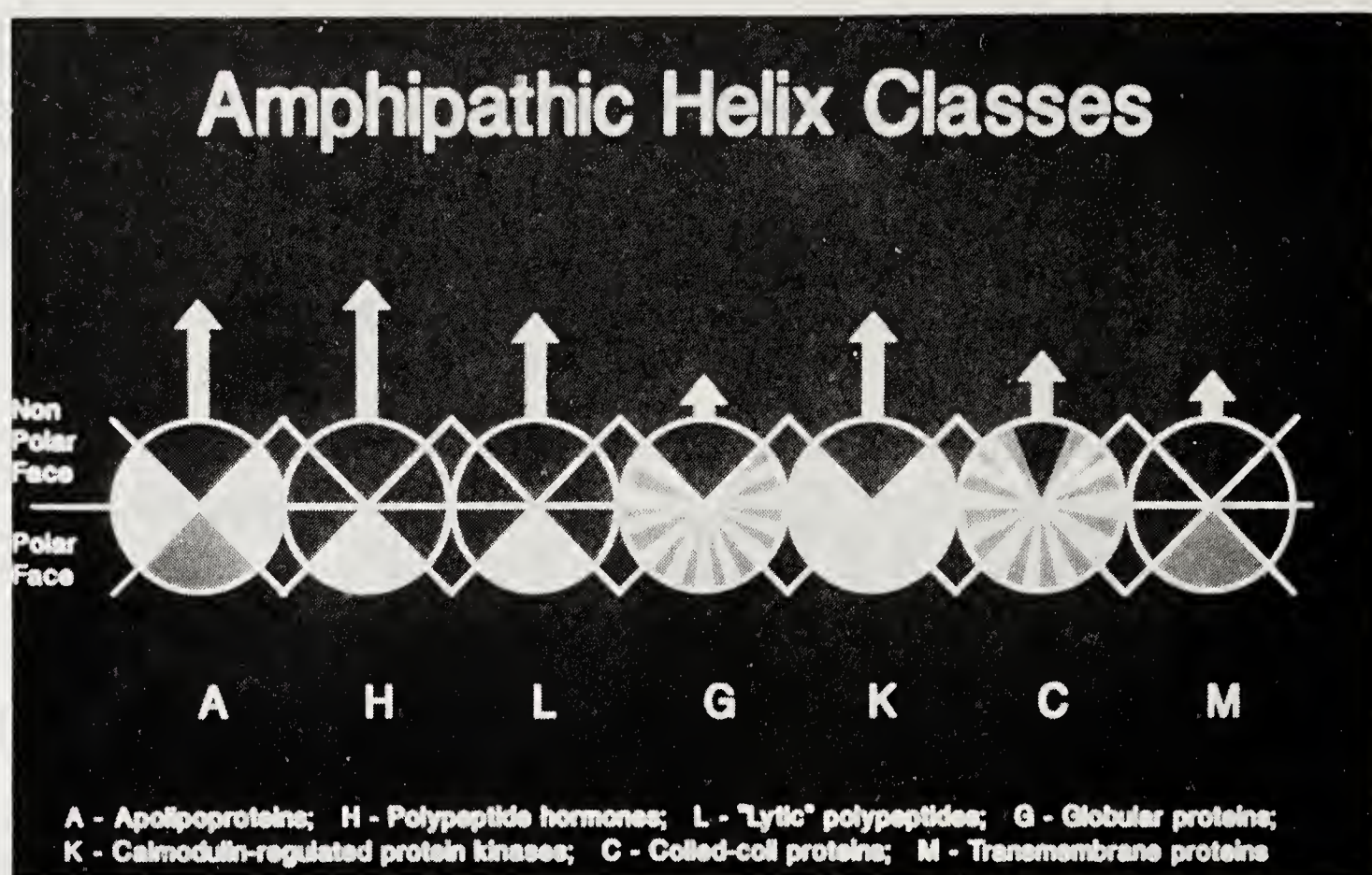


Figure 1. Schematic representation of seven different classes of the amphipathic helix. Each class is represented by a different circle. Arrows above each class show the relative mean hydrophobic moment. Black represents the nonpolar face; grey represents the polar face with acidic residues; and white, the polar face with basic residues.

important role in increasing the lipid affinity of the peptide analogs of the class A amphipathic helix (3--8).

The N-terminal amino acid residues in 18A and 18R are Asp and Lys, respectively; therefore it is possible that differences in helicity and lipid-affinity are due to the difference in the helix dipole moment between 18A and 18R. To address this question, in one series of experiments we have synthesized modified 18A and 18R analogs in which N- and C-terminal residues were blocked to obtain Ac-18A-NH₂ and Ac-18R-NH₂.

In a second series of experiments, analogs were synthesized in which we replaced a single residue with an amino acid having varying side chains. Three analogs were synthesized of the consensus 22mer peptide, a peptide whose sequence is based upon the consensus sequence for the eight 22mer tandem repeating units of apo A-I (3). In these analogs an amino acid on the nonpolar face but close to the polar-nonpolar interface (about 40° from the interface), was substituted by glutamic acid, homoaminoalanine (Haa) or lysine (figure 2b). Haa is a basic amino acid, but compared with lysine, possesses a shorter chain length, equal to that of glutamic acid. The peptide analog containing this amino acid thus acts as a control peptide to study the effect of chain length on lipid affinity. The three analog peptides synthesized with these substitutions at the 13th positions were designated [Lys¹³]A-Icon, [Glu¹³]A-Icon, and [Haa¹³]A-Icon.

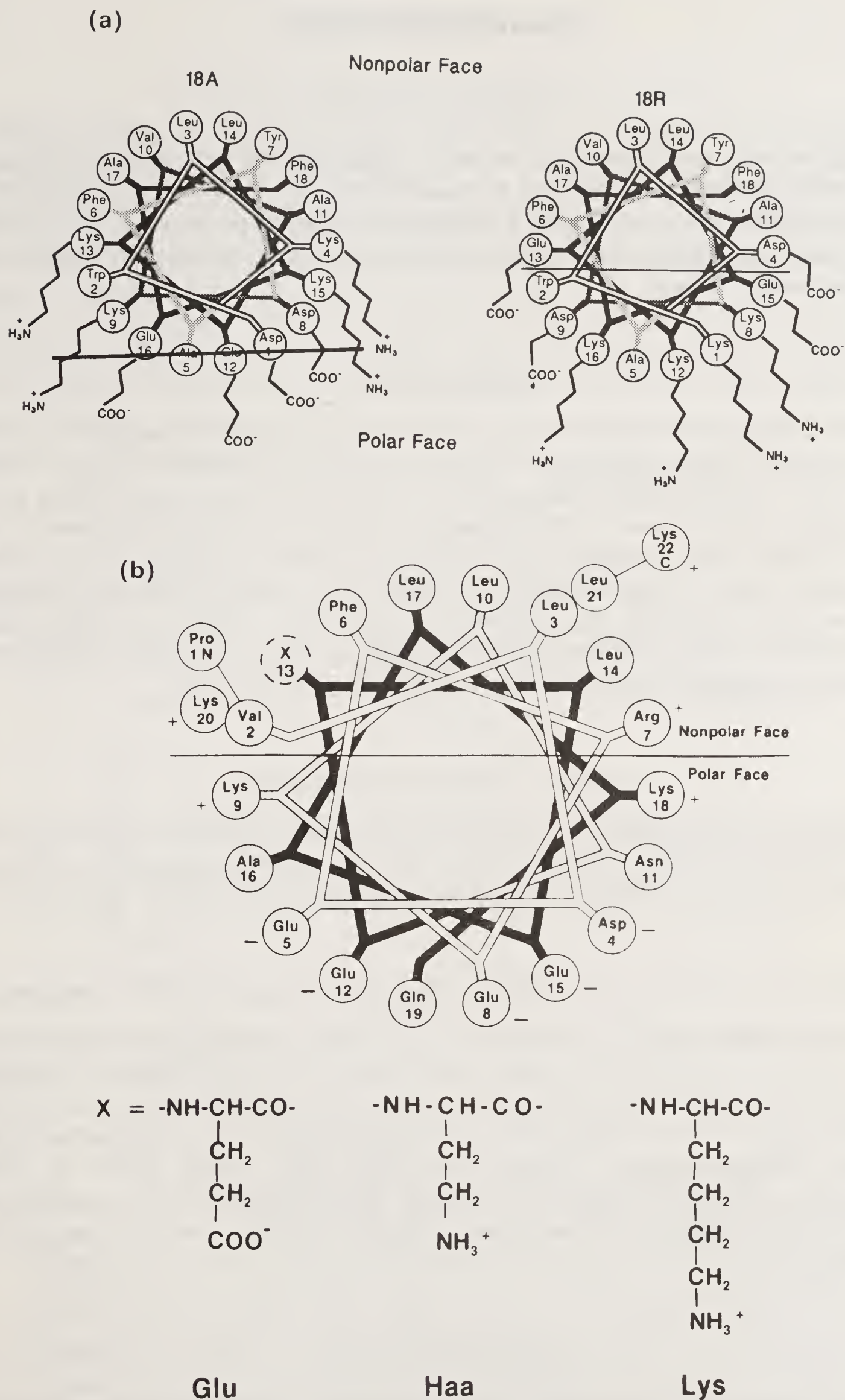


Figure 2. Helical wheel representation of (a) 18A and 18R; The line in 18A represents the hypothesis that the entire amphipathic helix is buried inside the membrane due to interfacial lysines. (b) A-I_{con} peptides with variation of the 13th residue in each of the peptides.

Materials and Methods

Synthesis of Peptides

Peptides were synthesized by the solid phase peptide synthesis procedure as previously described (3–5). Release of peptides from the solid support was achieved by the modified hydrogen fluoride procedure (4). Crude peptides were purified using high pressure liquid chromatography and characterized by amino acid analysis and sequence analysis (3–5).

Circular Dichroism

The CD spectra were taken using a JASCO spectropolarimeter coupled to a DPN 500 integrator. The peptides were dissolved in 10 mM phosphate buffer, 150 mM NaCl, pH 4. The peptides (500 μ g/ml) were mixed with lipid in the buffer at the lipid/peptide (M/M) ratio of 20:3% cholate was added to get clear solutions. These clear solutions were incubated at 4°C overnight. Then the solutions were subjected to extensive dialysis using 1000 MW cut off bags. The concentrations of the peptides were determined again by observing the optical densities at 280 nm in 4M guanidine hydrochloride. The concentrations of the peptides in the solutions of complexes were adjusted to 100 μ g/ml. The path length of the cell used was 1 mm.

Intrinsic Tryptophan Fluorescence

Fluorescence spectra were recorded at room temperature using an Aminco SPF-500 spectrofluorometer, with excitation at 287 nm for peptide alone and for discoidal particles of peptide with egg PC prepared using the cholate dialysis method (3).

Interaction of Peptides with Carboxyfluorescein-Entrapped Egg PC Liposomes

A dry thin film of egg PC was dispersed in Tris buffer containing carboxyfluorescein (CF) and was sonicated in ice for about one hour with 5-minute bursts of sonication to obtain small unilamellar vesicles. The solution was centrifuged to free the metallic particles. Free CF was removed by gel chromatography on a Sepharose CL-4B column. CF-encapsulated vesicles were mixed with Triton X-100 to obtain fluorescence reading for complete dye leakage. Known amounts of the peptides were added and the change in the fluorescence intensity was monitored for a period of 25 minutes.

Electron Microscopy

The peptides were complexed with unilamellar vesicles of dimyristoylphosphatidylcholine (DMPC) at 1:1 weight ratio, stained with 2% potassium phosphotungstate, pH 5.9, and examined with a Philips EM400 microscope on carbon coated Formvar grids.

Results

Circular Dichroism and Tryptophan Fluorescence

Peptide analogs Ac-18A-NH₂ and Ac-18R-NH₂ were subjected to CD studies in aqueous buffer alone or in the presence of DMPC. The results are shown in figure 3. For comparison, the CD studies of 18A are also shown. The results indicate that the increase in helicity in the presence of lipid is larger for Ac-18A-NH₂ than for Ac-18R-NH₂. Consistent with these results are the intrinsic tryptophan fluorescence studies of Ac-18A-NH₂ which showed a blue shift (from max 358 nm in buffer to 338 nm in presence of DMPC) indicating the transfer of Trp from the polar environment to the nonpolar environment. There was no blue shift for Ac-18R-NH₂ in the presence of lipid. Further, while 18A and Ac-18R-NH₂ exhibit similar CD

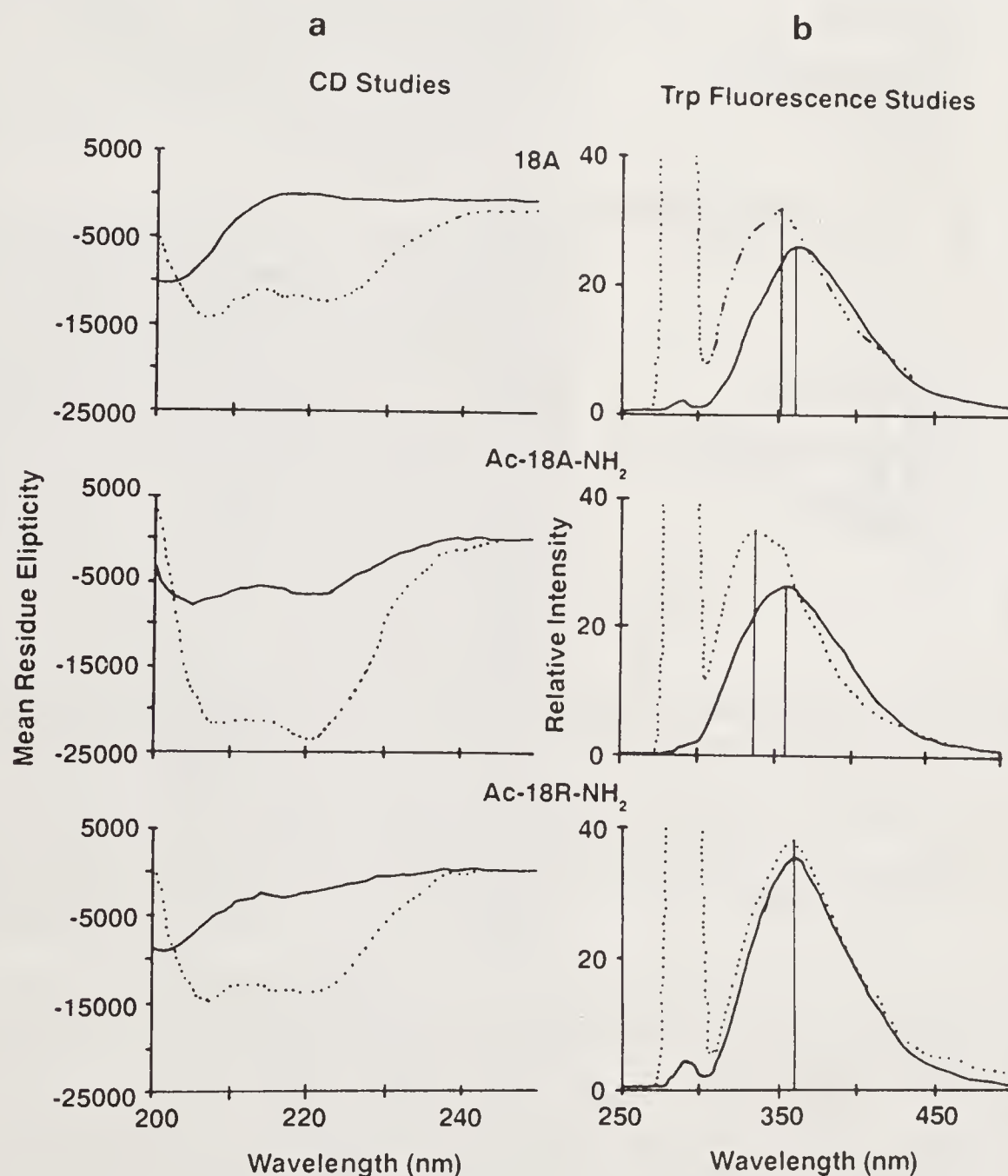


Figure 3. (a) Circular dichroism spectra of 18A (top), Ac-18A-NH₂ (middle), and Ac-18R-NH₂ (bottom). The solid lines represent peptides in phosphate buffer at pH 7.4, and the dashed lines represent peptides in presence of DMPC at a lipid/peptide (M/M) ratio of 20. The peptide concentrations were maintained at 50 μ g/ml. (b) Intrinsic tryptophan fluorescence spectra of the peptides (50 μ g/ml) 18A (top), Ac-18A-NH₂ (middle), and Ac-18R-NH₂ (bottom). The solid lines represent the emission spectra for the peptides in phosphate buffer at pH 7.4, and dashed lines represent peptides with egg PC at lipid/peptide (M/M) ratio of 20. The solutions were excited at 287 nm.

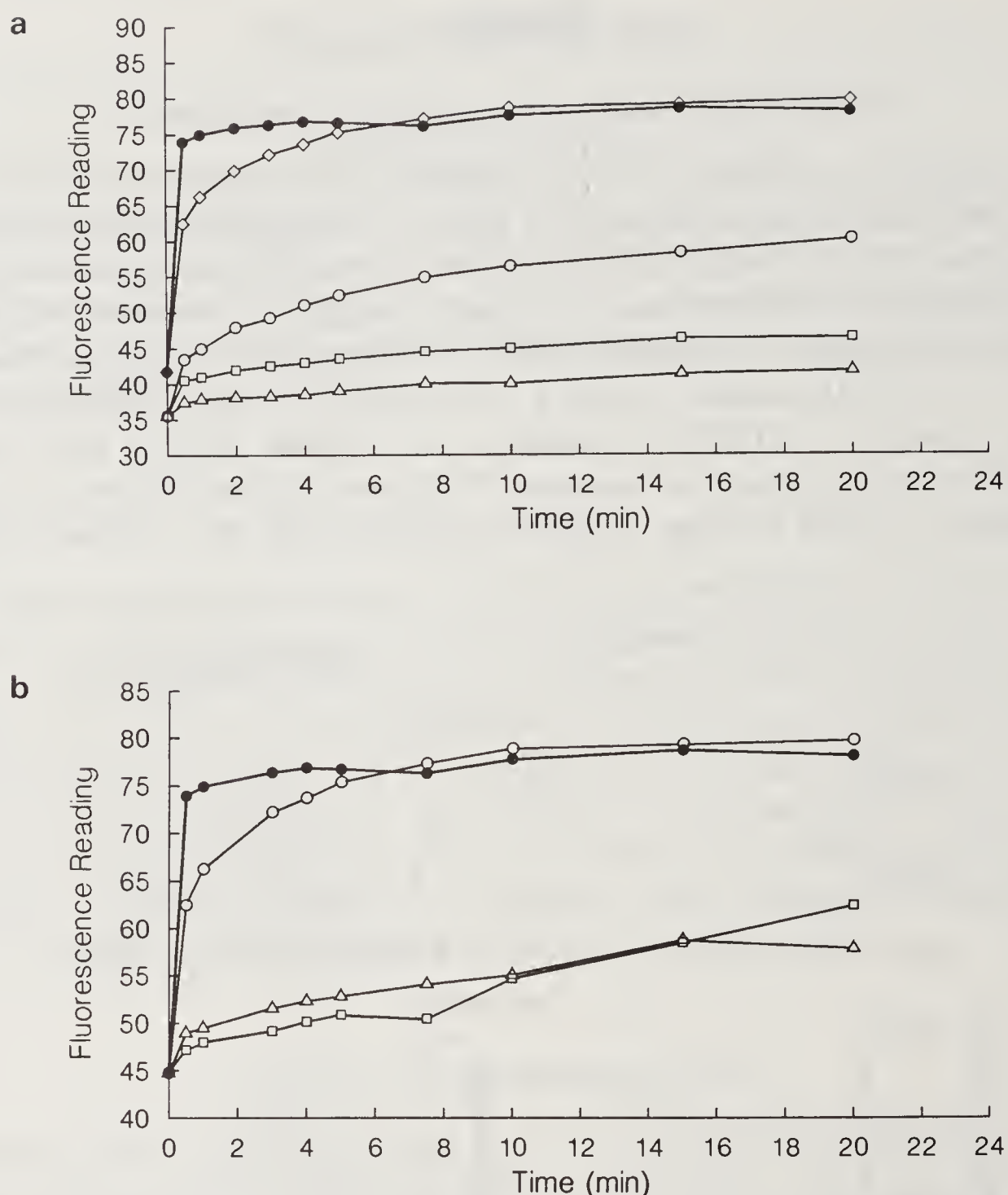


Figure 4. Carboxyfluorescein leakage from egg PC entrapped vesicles: (a) Different amounts of [Lys¹³]A-Icon analog were used to leak out the dye. (\triangle — \triangle), 25 μ g; (\square — \square), 50 μ g; (\circ — \circ), 100 μ g; (\diamond — \diamond), 200 μ g and (\bullet — \bullet), Triton X-100 (10 μ g). (b) Different peptides (200 μ g) and Triton X-100 (10 μ g): (\square — \square), [Haa¹³]A-Icon; (\triangle — \triangle), [Glu¹³]A-Icon; (\circ — \circ), [Lys¹³]A-Icon and (\bullet — \bullet), Triton X-100.

pattern both in aqueous buffer and in the presence of lipid, 18A does show considerable shift in the tryptophan maximum indicating a greater lipid affinity for 18A compared to Ac-18R-NH₂. Right-angled light scattering studies (see ref. 7) of 18A and Ac-18R-NH₂ with DMPC showed that while 18A clarified multilamellar vesicles of DMPC at 1:5 and 1:10 (peptide:lipid) molar ratio, Ac-18R-NH₂ did not, thus supporting the results from tryptophan fluorescence studies.

Interaction of Peptides with CF-Entrapped Egg PC Small Unilamellar Liposomes

Two experiments were carried out. In the first, different amounts of [Lys¹³]A-Icon were interacted with a constant amount of dye-entrapped egg PC to determine the minimum concentration at which this peptide equals that of Triton X-100 in its ability to produce complete dye leakage (figure 4a). In the second experiment, the other two peptide analogs were studied at the minimum concentration determined for

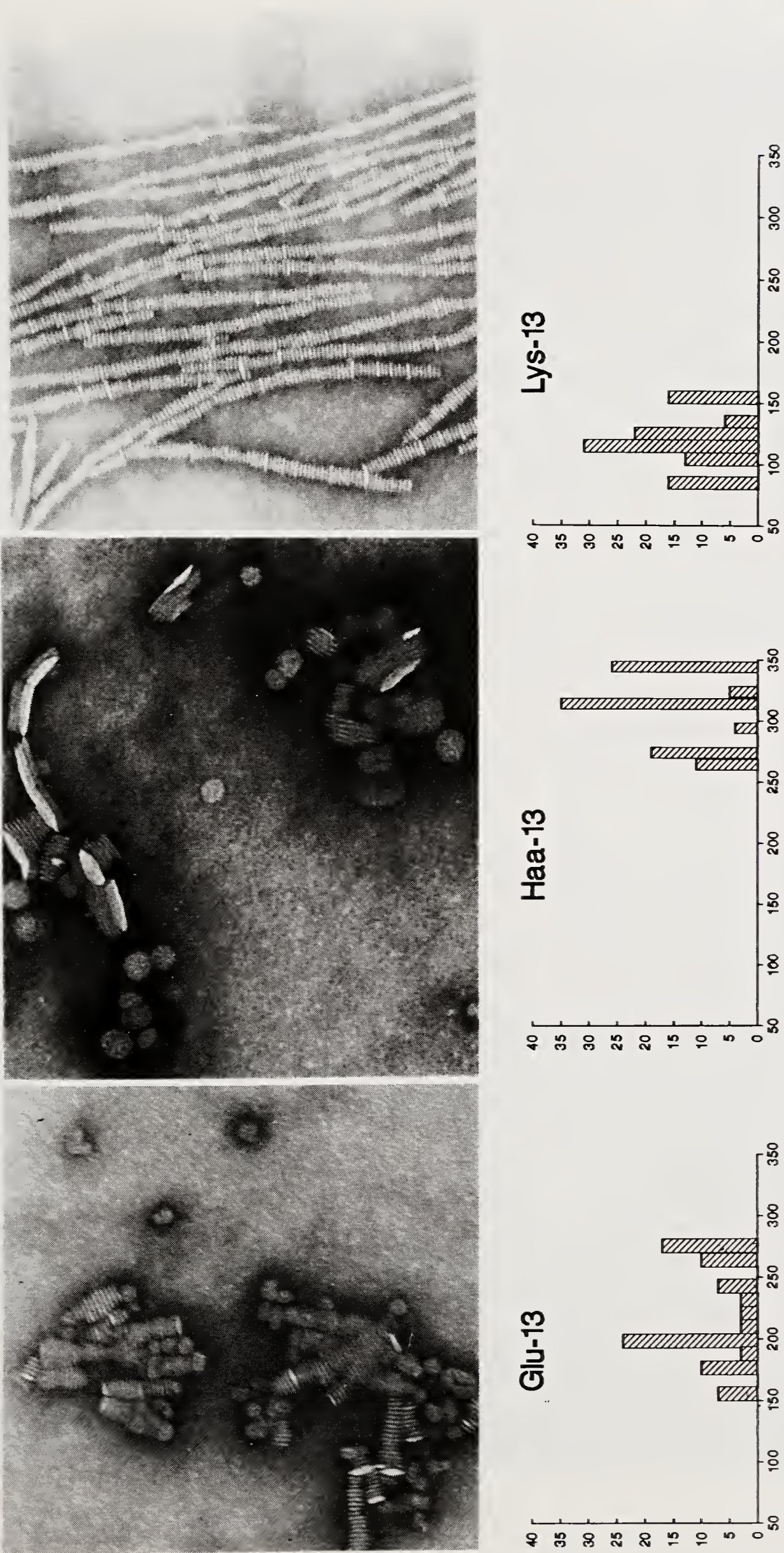


Figure 5. Negative stain electron microscopy of peptide DMPC complexes at 1:1 weight ratios; and morphographic analysis of diameters of discoidal complexes.

[Lys¹³]A-Icon (figure 4b). The results show that [Lys¹³]A-Icon at 200 μ g equals Triton X-100, whereas, the peptide analogs [Glu¹³]A-Icon and [Haa¹³]A-Icon, at the same concentration, released only about 50% of the entrapped dye.

Electron Microscopy

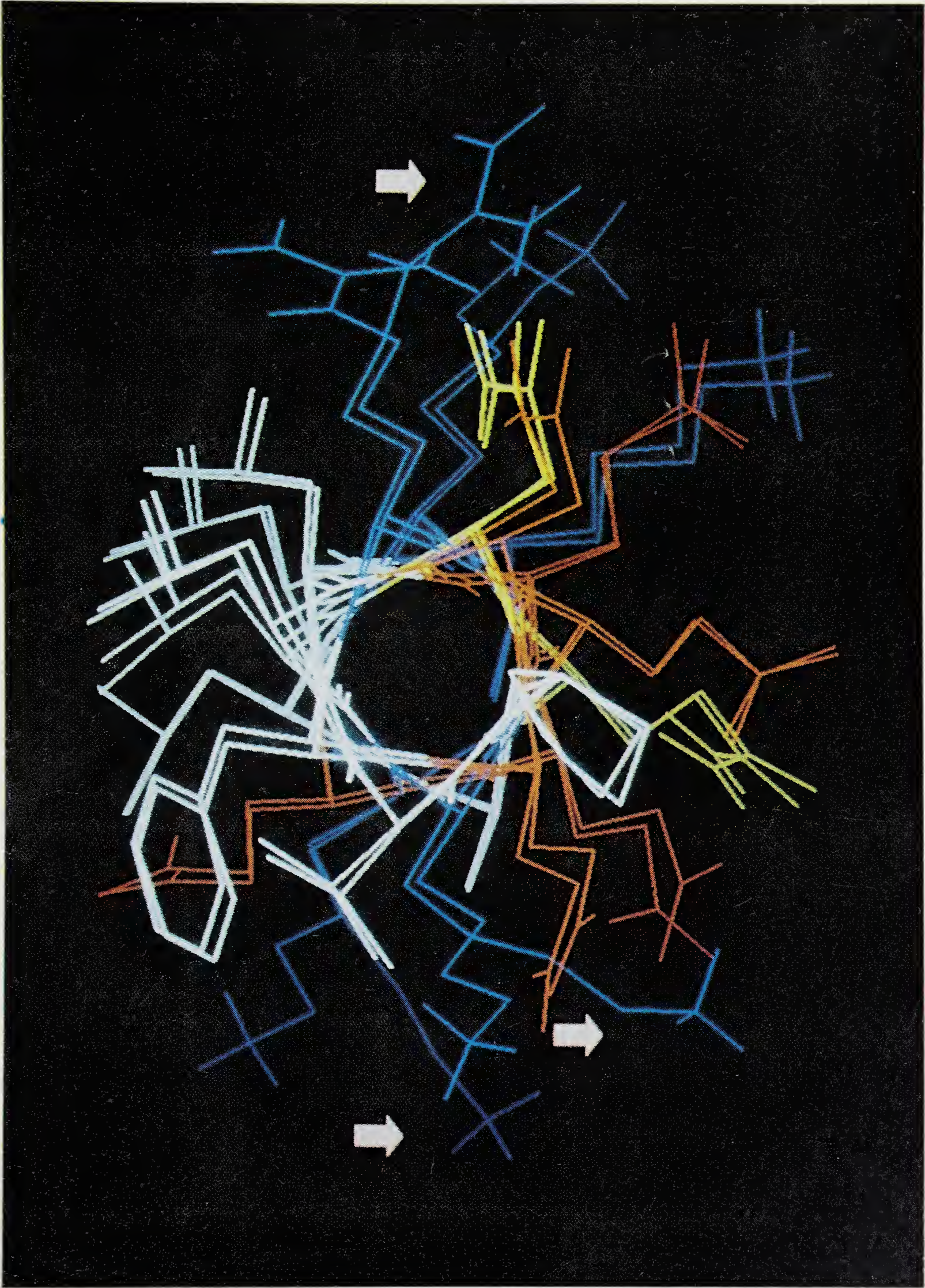
Leakage of CF from the liposomes could be partially a kinetic phenomenon and may not be a function entirely related to the lipid affinity of peptides. We therefore analyzed the particles formed by the interaction of these peptide analogs with DMPC. Electron micrographs of peptide:DMPC complexes supported the conclusions derived from previous experiments (figure 5). At a 1:1 weight ratio of peptide:lipid, the [Lys¹³]A-Icon formed the smallest discoidal complexes of the three (Stoke's diameters: [Lys¹³]A-Icon, 110 ± 30 Å; [Glu¹³]A-Icon, 200 ± 50 Å; [Haa¹³]A-Icon, 310 ± 30 Å). We have proposed elsewhere (3,4) that disc diameters are inversely proportional to the lipid affinity of synthetic peptides. These results indicate a rank order for lipid affinity of [Lys¹³]A-Icon > [Glu¹³]A-Icon \geq [Haa¹³]A-Icon.

Discussion

With the discovery of the amphipathic helix as a structural motif that mediates the association of plasma apolipoproteins with lipids, it was noted that the amphipathic helices of apolipoproteins possess a unique charge distribution (1). Positively charged residues are localized at the polar–nonpolar interface and negatively charged residues at the center of the polar face in the class A amphipathic helix. Our earlier peptide analogs were designed to understand the significance of such a charge distribution, and we found that the peptide analogs with a reversed-charge distribution were less effective in their lipid-binding ability (5,6).

These results led to a careful examination of lipid-associating amphipathic helices present in other biologically active peptides and proteins (2). The presence of positively charged residues at the polar–nonpolar interface is apparently unique to amphipathic helices found in apolipoproteins (including such diverse apolipoproteins as serum amyloid A). However, the amphipathic helices associated with certain short polypeptide hormones (H), lytic peptides (L), and calmodulin-regulated protein kinases (K) are characterized by the presence of a high density of positively charged residues. To investigate these unique features of amphipathic helices and to understand the role played by positively charged amino acids in the different classes of amphipathic helices, we have initiated a study of model synthetic peptide analogs representing these different classes.

Figure 6. Computer model of the consensus amphipathic peptide. The helix is viewed down its axis. Non-polar residues are represented in white, the polar face is down. The features of the "snorkel model" are indicated by the arrows. Side chain dihedral angles of each of the positively charged residues are changed from *trans* to *gauche* conformation so as to bring the ends of the side chains closer to the polar side of the helix. The energy difference between the conformations of the two structures is very small, showing that the snorkel configuration is possible without intramolecular steric constraints.





Elephants
are wonderful
swimmers, in spite of their
great weight and size. They
often swim with their bodies
almost completely under
the water, with only
their trunks sticking
out. They can snor-
kel like this for
hours, and swim
long distances.

Figure 7. A cartoon showing elephants swimming in water with their trunks SNORKELED above the water. (Reproduced with permission from Zoo Books, San Diego, CA.)

We attribute reasons for the higher lipid affinity of 18A when compared to 18R to three factors:

- 1) The interfacial Lys in 18A contributes to the overall increased hydrophobicity of the amphipathic helix, and thus contributes to its increased lipid affinity;
- 2) The position of charged residues are reversed in the two peptides resulting in different helical dipole moments; and,
- 3) 18R does not have the potential to form an α -helical structure, and thus has a weaker lipid affinity.

To increase the α -helicity and to attempt to make the helical dipole moments similar, we synthesized analogs with capped end groups, Ac-18A-NH₂ and Ac-18R-NH₂. These modifications caused an increase in the helicity for both the peptides. 18A and Ac-18R-NH₂ both had about 15% α -helicity in aqueous buffer alone and 40% helicity in the presence of lipids. On the other hand, Ac-18A-NH₂ was approximately 50% helical in buffer alone and approximately 65% helical in the presence of lipids. Tryptophan fluorescence studies show that, although the Ac-18R-NH₂ was as helical as the parent 18A, there was no measurable tryptophan blue shift.

In our design of model peptide analogs, Trp is present at the polar face, close to the polar–nonpolar interface. We favor the hypothesis that the results of the tryptophan fluorescence studies in the presence of lipid are reflective of the depth of burial of the amphipathic helix in the lipid bilayer. We postulate that the increased Trp-fluorescence shift observed for Ac-18A-NH₂ compared with Ac-18R-NH₂ is due to the participation of the alkyl side chains of interfacial Lys residues. Thus, essentially the entirety of the Ac-18A-NH₂ amphipathic helix is buried within the hydrophobic face, making Trp and lipid acyl chains interaction more favourable as shown in figure 2a. Thus, in this series of peptide analogs, the extent of shift in the Trp fluorescence maximum is an indicator of the depth of burial of the amphipathic helical peptide and is also approximately proportional to the lipid affinity. This conclusion is consistent with the turbidity clarification studies (results not shown).

The higher lipid affinity of 18A compared with 18R is supported by two previous experiments from our laboratory:

- 1) Increasing the α -helicity of 18R peptides by substituting amino acids with high helix-forming potential resulted in 18R analogs which possessed a weaker lipid affinity than the corresponding 18A analogs (7,8).
- 2) We demonstrated a rank order of lipid affinity 18A > 18R when the ends were capped by either three serine or three alanine residues (9).

We therefore conclude that 18A and Ac-18A-NH₂ have higher lipid affinity than Ac-18R-NH₂ in a manner that is independent of the effect of the helix dipole.

To study the effect of alkyl chain length of interfacial lysines (and arginines) on the lipid affinity of the amphipathic helix, we designed three peptide analogs of the apo A-I consensus peptide for the eight 22-mer lipid-associating domains of apo A-I (figure 1b). This sequence was constructed by the selection of the most prevalent amino acid residues in each position of different tandem repeating 22mer units of apo A-I. In these peptides, single amino acid substitutions were made at the 13th residue (figure 2b) using lysine, glutamic acid and homoaminoalanine. The 13th position is on the nonpolar face but is close to the polar–nonpolar interface of the amphipathic helix. Because both the amino acids, lysine and homoaminoalanine,

possess a side-chain amino group, the peptides derived using these two amino acids at the 13th position should have the similar helix dipoles, and the lipid affinity in these peptides should therefore be a function of the difference in the alkyl chain between these two amino acids (two-CH₂-in Haa compared to four-CH₂-in Lys). Further, the peptide derived from homoaminoalanine serves as a control peptide, because it has the same acyl chain length as glutamic acid but with a similar side chain functionality to lysine.

Studies from these peptides convincingly demonstrated the following rank-order of lipid affinity for these peptides: [Lys¹³]A-Icon \gg [Glu¹³]A-Icon \geq [Haa¹³]A-Icon. Based on these results we proposed the "snorkel" model, shown in figure 6, for the lipid association of the apolipoprotein class of amphipathic helices. The bulk of the van der Waal's surface areas of the positively charged residues are hydrophobic. We propose that the amphipathic basic residues, when associated with phospholipid, extend toward the polar face of the helix to insert their charged moieties into the aqueous milieu for solvation. The concept is diagrammatically illustrated in figure 7, where elephants are able to swim because of the ability of their trunks to snorkel out of water beyond the air-water interface. As the result of basic residues snorkeling, we suggest that essentially all of the uncharged van der Waal's surface of the amphipathic helical domains of the apolipoproteins can be buried within the interior of the phospholipid monolayer, thus providing for a deeper helix insertion as well as an increased lipid affinity.

The snorkel model may be relevant to some of the differences between the three classes of lipid-associating amphipathic helices. The amphipathic helices of classes H and L have narrow positively charged amino acid residues clustered toward the center of the polar face. The snorkel model predicts that this particular amphipathic helix organization would allow deeper penetration into a membrane bilayer than an amphipathic helix with either a broader face or a zwitterionic face.

Deep membrane penetration might be important in the biological effects of classes H and L. The interaction of polypeptide hormones with receptors may involve a deep penetration of their amphipathic helical domains to produce contact with an intramembraneous portion of the receptor or to induce some coordinate alteration in the properties of the inner half of the membrane bilayer. Lysyl residues have a longer hydrophobic side chain than arginyl residues; the restriction of the positive residues in the "lytic" class of amphipathic helices to lysyl residues may allow deeper membrane penetration and therefore, the destabilization of membrane integrity. Thus, it may be that certain classes of amphipathic helices can produce signal transduction via direct interaction with the cell membrane, bypassing cell surface receptor interactions (10). We believe that further studies of model peptide analogs corresponding to different classes of the amphipathic helix may provide information that will aid in understanding the biological roles assumed by the different classes of amphipathic helix.

Acknowledgements

This work was supported by National Institute of Heart, Lung and Blood Grants HL-34343.

References

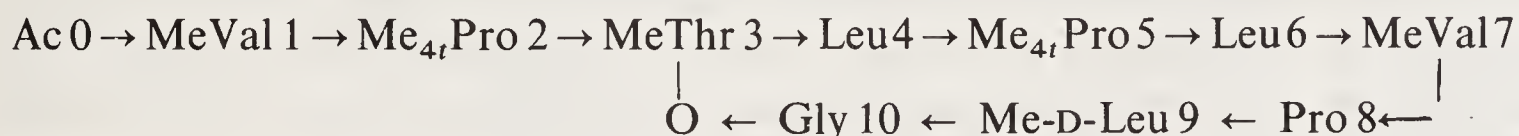
1. Segrest, J. P., Jackson, R. L., Morrisett, J. D. & Gotto, A. M., Jr. (1973) *FEBS. Lett.* **38**, 247–253.
2. Segrest, J. P., DeLoof, H., Dohlman, J. G., Brouillette, C. G. & Anantharamaiah, G. M. (1990) *Proteins* **8**, 103–117.
3. Anantharamaiah, G. M., Venkatachalapathi, Y. V., Brouillette, C. G. & Segrest, J. P. (1990) *Arteriosclerosis* **10**, 95–105.
4. Anantharamaiah, G. M. (1985) *Methods Enzymol.* **128**, 628–668.
5. Anantharamaiah, G. M., Jones, J. L., Brouillette, C. G., Schmidt, C. F., Chung, B. H., Hughes, T. A., Bhowan, A. S. & Segrest, J. P. (1985) *J. Biol. Chem.* **260**, 10248–10255.
6. Chung, B. H., Anantharamaiah, G. M., Brouillette, C. G., Nishida, T. & Segrest, J. P. (1985) *J. Biol. Chem.* **260**, 10256–10262.
7. Epand, R. M., Gawish, A., Iqbal, M., Gupta, K. G., Chen, C. H., Segrest, J. P. & Anantharamaiah, G. M. (1987) *J. Biol. Chem.* **262**, 9389–9396.
8. Epand, R. M., Surewicz, W. K., Hughes, D. W., Mantsch, H., Segrest, J. P., Allen, T. M. & Anantharamaiah, G. M. (1989) *J. Biol. Chem.* **264**, 4628–4635.
9. Segrest, J. P., Chung, B. H., Brouillette, C. G., Kanellis, P. & McGahan, R. (1983) *J. Biol. Chem.* **258**, 2290–2295.
10. Jorgenson, E. V., Anantharamaiah, G. M., Segrest, J. P., Gwynne, J. W. & Handwerger, S. (1989) *J. Biol. Chem.* **264**, 9215–9219.

2D-NMR Conformational Analysis of Griselimycin, an Antituberculous Cyclodepsipeptide

Manh Thong Cung, Bernard Vitoux,
Pascal Demange and Michel Marraud

Laboratoire de Chimie Physique Macromoléculaire, CNRS-URA-494,
ENSIC-INPL, BP 451, 54001 Nancy Cedex, France.

Griselimycin, isolated from *Streptomyces coelicus* and *griseus* (1), is a lactonic cyclodepsipeptide with the following primary structure:



It exhibits an important antimicrobial activity, especially against tuberculosis. Like many antibiotic peptides extracted from plants (2–5) or micro-organisms (6–13), griselimycin contains various structural characteristics: a cyclization (2–13) and inclusion of a lactonic bond (6–13), non-coded N-methylamino acids (2–13) and a D-amino acid (6, 7, 13).

A number of cyclic and acyclic analogues of griselimycin have been synthesized by Stverteczky and Bajusz (14) to elucidate the structure–activity relationship of this compound. As for most macrocyclic peptide antibiotics (virginiamycin, etamycin, actinomycin...) (7, 8, 10), the acyclic analogues of griselimycin are devoid of antibacterial activity. On the contrary, the cyclic analogues exhibit biological activity, but only when the MeLeu 9 residue has the D-configuration, even when one amino acid residue is deleted in the cycle. The local structure involving the Me-D-Leu 9 residue seems to have a particular importance for the biological expression of griselimycin. Therefore, the conformational description of this antibiotic in solution is of interest for the study of its structure–activity relationship.

Due to the hydrophobic character of nearly all the side-chains, griselimycin is insoluble in aqueous solution. In the present study, we have determined its conformation in chloroform and dimethylsulfoxide (DMSO) by one- and two-dimensional ^1H and ^{13}C -NMR spectroscopy. The ^1H - and ^{13}C -NMR spectra have been completely assigned by combining ^1H , ^1H -COSY, ^1H , ^{13}C -COSY, relayed COSY, HOHAHA and NOESY experiments. A conformational analysis of the cyclodepsipeptide has been carried out through measurements of:

- (i) the homo (^1H , ^1H) and heteronuclear (^1H , ^{13}C) vicinal coupling constants,
- (ii) the temperature and solvent dependence of the peptide NH proton chemical shifts and
- (iii) the interproton distances estimated from nuclear Overhauser effects.

Results and Discussion

NMR Spectral Assignment

On the basis of the conformational dependence of the NMR parameters, the unambiguous assignment of the chemical shifts and coupling constants is a prerequisite for the conformational analysis of peptides in solution.

^1H resonances: The 1D-NMR spectrum of griselimycin in CDCl_3 and $\text{DMSO}-d_6$ (figure 1) reveals no *cis-trans* equilibrium in spite of the presence of several tertiary amide bonds. However, the high field part of the spectrum (0.6–2.5 ppm) which contains the side-chain proton resonances is highly crowded and must be assigned by 2D-NMR techniques.

Usually, the first step for the analysis of peptide NMR spectra is the identification of the connectivities within the proton spin systems for each amino-acid residue by ^1H , ^1H -COSY experiments. However some uncertainties can be due to the overlapping signals, especially in the high field region containing the aliphatic resonances. In this case, relayed COSY spectroscopy which includes a second magnetization transfer step reveals the protons which are indirectly coupled via a common coupling partner, in addition to the classical cross-peaks observed with the COSY spectrum. For example, figure 2 shows the four bond connectivities between all the α and γ protons of griselimycin (circled cross-peaks).

The distinction between the identical NMe-Val 1/NMe-Val 7, Me_{4t}-Pro 2/Me_{4t}-Pro 5 and Leu 4/Leu 6 spin systems, and the assignment of the four N-methyl groups cannot be obtained via the conventional COSY techniques, but rather from the NOE effects between the NH or N-Me protons and the α -proton of the neighbouring residues (15). For example, the Leu 4 residue is characterized by two NOESY

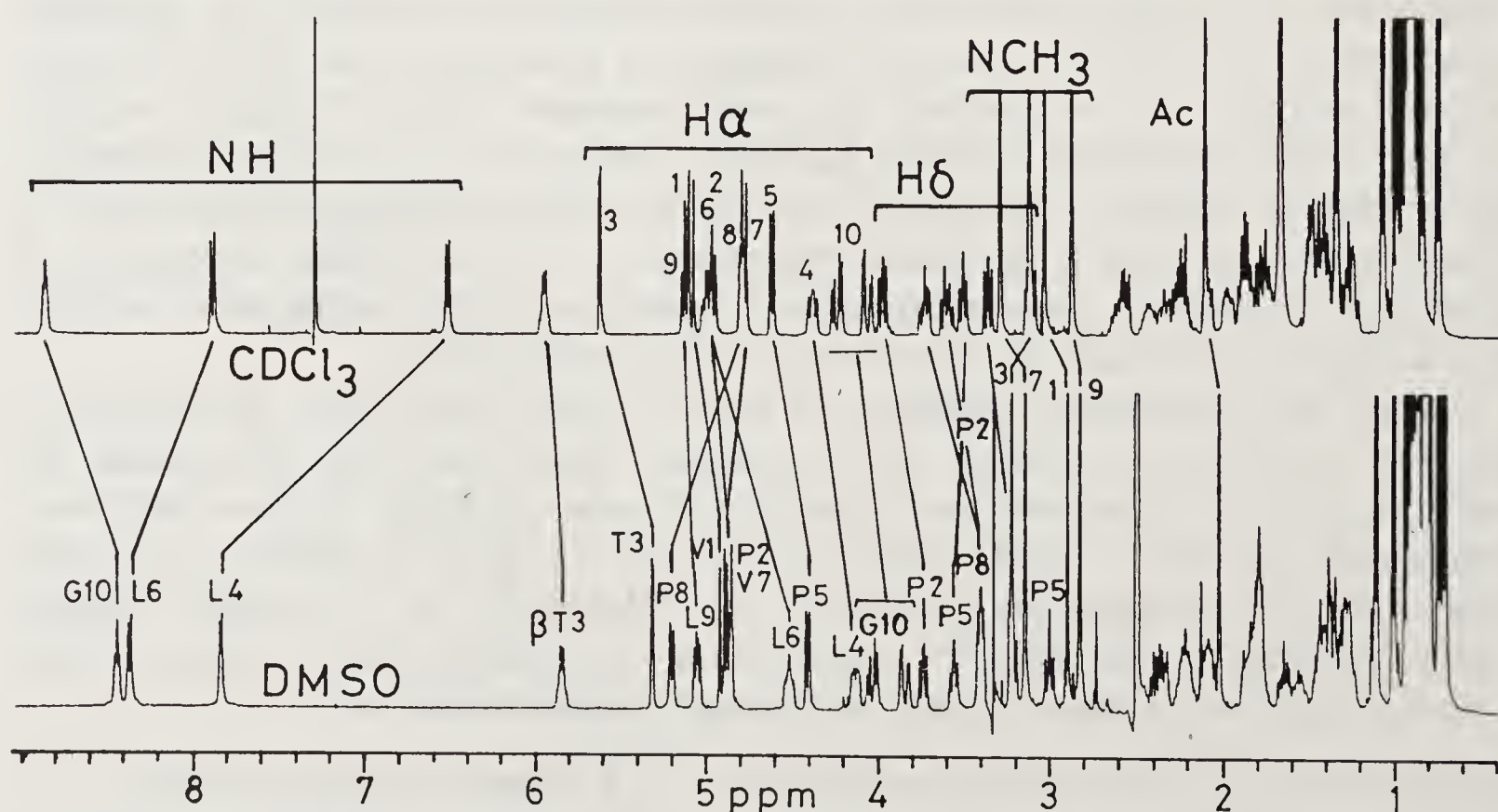


Figure 1. One-dimensional 400MHz ^1H -NMR spectra of griselimycin in chloroform and dimethylsulfoxide at 25°C. The assignment of most of the protons is specified.

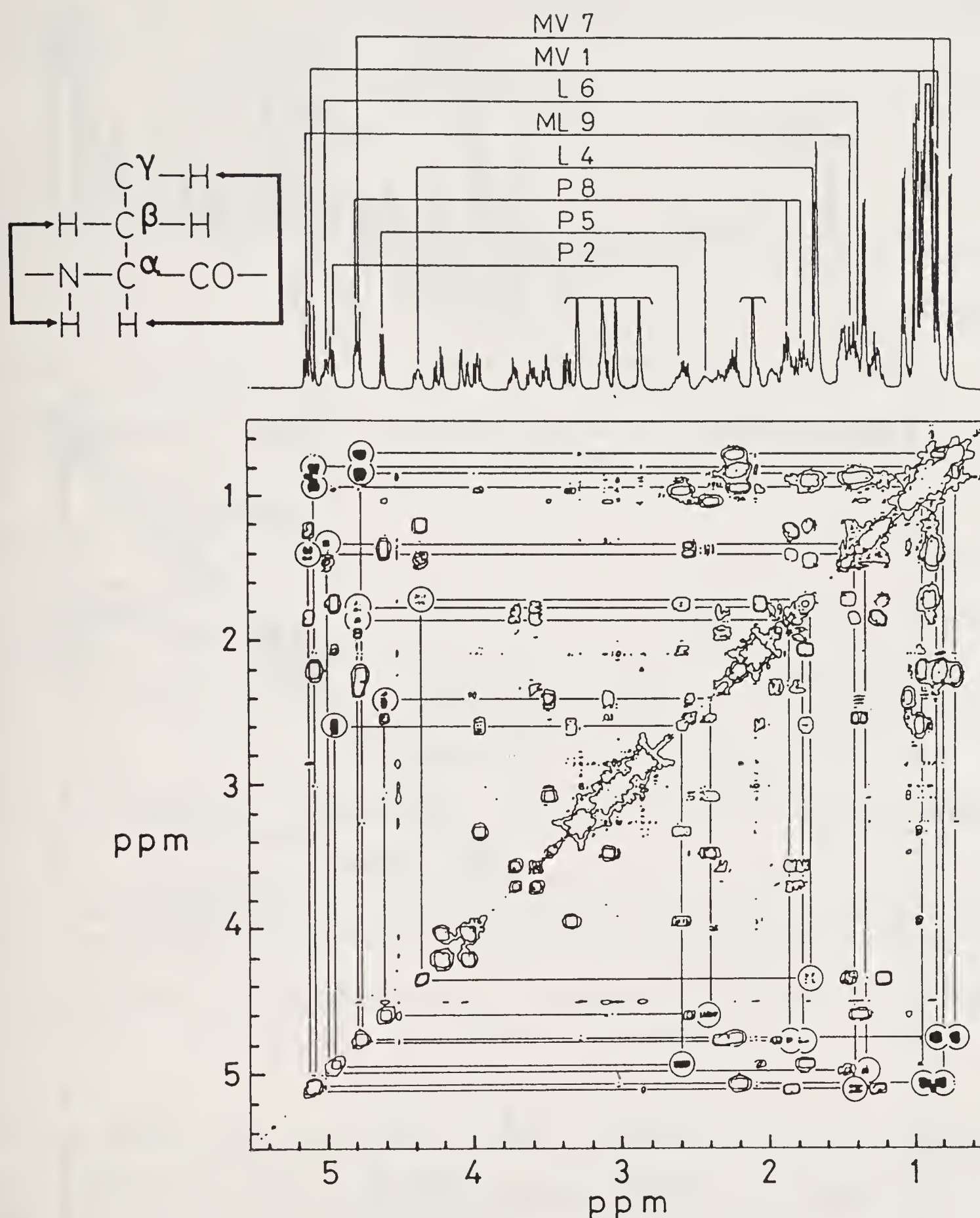


Figure 2. Aliphatic portion of the relayed ^1H , ^1H -COSY spectrum of griselimycin in chloroform. The circled cross-peaks indicate the α/γ -proton connectivities.

cross-peaks between its NH signal at 7.8 ppm and the NMe-L-Thr 3 α and β -proton resonances at 5.3 and 5.8 ppm, respectively. In the same way the Leu 6-NH signal is correlated with the Me_{4t}-Pro 5 α -proton (figure 3).

The four N-methyl groups have been attributed by considering their connectivity with the α -proton of the preceding residue. Effectively, cross-peaks are observed between Val 1-NMe and the acetyl methyl group, Thr 3-NMe and the Pro 2 α -proton, Val 7-NMe and the Leu 6 α -proton, and finally D-Leu 9-NMe and the Pro 8 α -proton. Moreover, the NOESY cross-peaks also discriminate the NMe-Val 1/NMe-Val 7 and Me_{4t}-Pro 2/Me_{4t}-Pro 5 residues, and allow the sequential assignment of all the

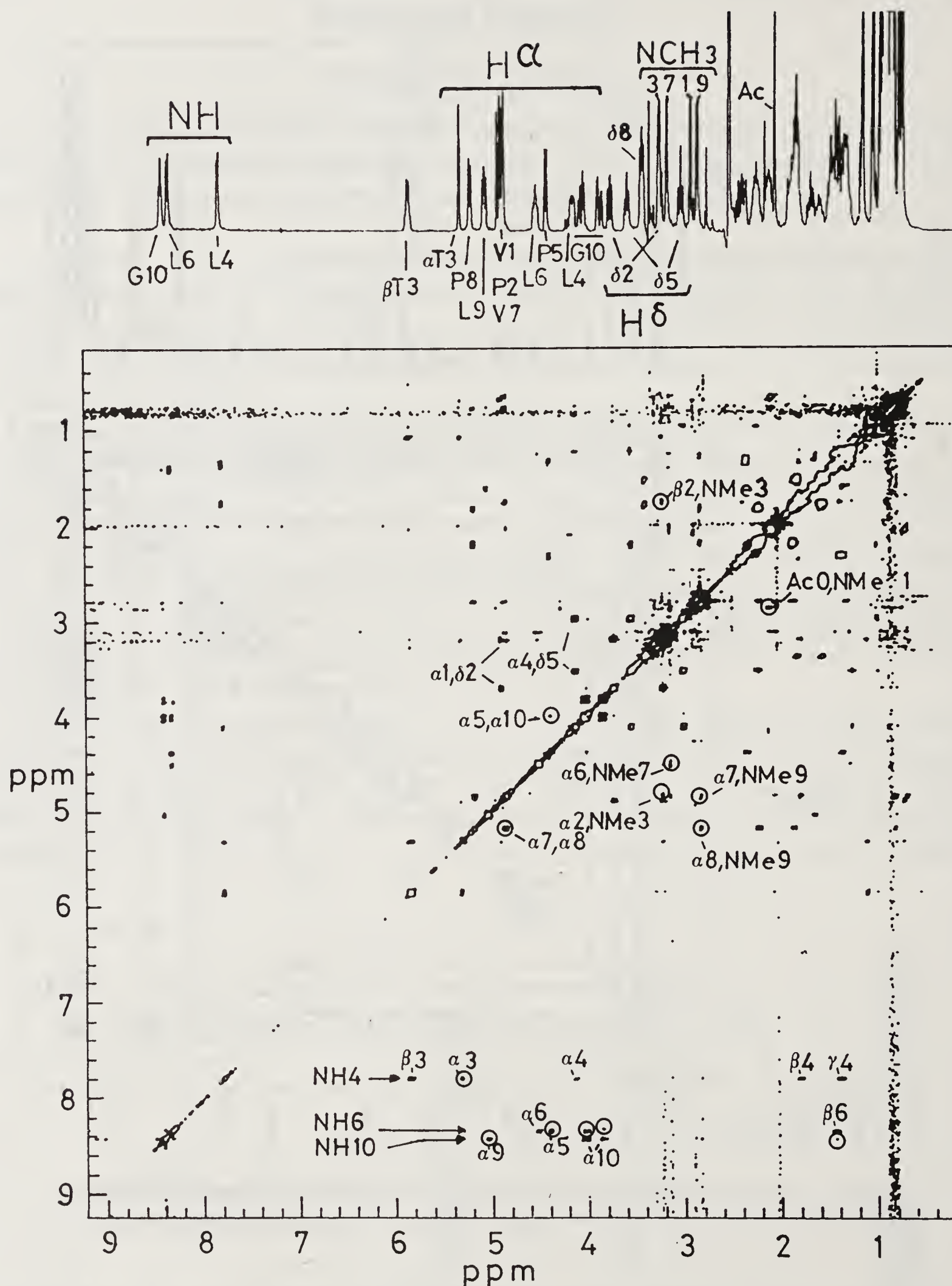


Figure 3. NOESY spectrum of griselimycin in dimethylsulfoxide at 25°C (mixing time: 300 ms). The main cross-peaks are attributed.

resonances. The most striking fact is the presence of a strong correlation between the Val 7 and Pro 8 α -protons indicating a *cis* conformation of the Val 7–Pro 8 amide bond.

The chemical shifts of the proton resonances in CDCl_3 and $\text{DMSO}-d_6$ are given in table 1.

Table 1. ^1H -NMR chemical shifts (in ppm from TMS) of griselimycin resonances in CDCl_3 and $\text{DMSO}-d_6$.

	NH	NCH_3	C^αH	C^βH^a	C^γH	C^δH	others
<i>CDCl₃</i>							
CH_3CO							2.12 (Ac)
NMe-L-Val 1	—	3.04	5.07	2.21	0.99; 0.85		
Me_{4t} -L-Pro 2	—	—	4.94	2.09 (c); 1.76 (t)	2.60	3.98; 3.36	1.01 (Me_{4t})
NMe-L-Thr 3	—	3.12	5.61	5.94	1.35		
L-Leu 4	6.49	—	4.37	1.46 (R); 1.23 (S)	1.75	0.95; 0.93	1.08 (Me_{4t})
Me_{4t} -L-Pro 5	—	—	4.61	2.55 (c); 1.39 (t)	2.43	3.51; 3.11	
L-Leu 6	7.85	—	4.99	1.51 (S); 1.50 (R)	1.35	0.92; 0.90	
NMe-L-Val 7	—	3.29	4.77	2.26	0.89; 0.76		
L-Pro 8	—	—	4.78	2.34 (c); 1.96 (t)	1.88; 1.79	3.73; 3.60	
NMe-D-Leu 9	—	2.88	5.11	1.88 ^b ; 1.28 ^b	1.41	0.93; 0.92	
Gly 10	8.83	—	4.23; 4.06				
<i>DMSO-d_6</i>							
CH_3CO							2.03 (Ac)
NMe-L-Val 1	—	2.90	4.89	2.11	0.89; 0.75		
Me_{4t} -L-Pro 2	—	—	4.84	1.79; 1.79	2.39	3.73; 3.21	0.94 (Me_{4t})
NMe-L-Thr 3	—	3.22	5.29	5.83	1.13		
L-Leu 4	7.82	—	4.12	1.40 (R); 1.31 (S)	1.81	0.92; 0.90	
Me_{4t} -L-Pro 5	—	—	4.40	2.34 (c); 1.36 (t)	2.23	3.56; 3.00	1.02 (Me_{4t})
L-Leu 6	8.33	—	4.50	1.44 (S); 1.32 (R)	1.36	0.87; 0.80	
NMe-L-Val 7	—	3.14	4.85	2.07	0.84; 0.72		
L-Pro 8	—	—	5.19	2.21 (c); 1.86 (t)	1.81; 1.58	3.39; 3.41	
NMe-D-Leu 9	—	2.83	5.03	1.64 ^b ; 1.35 ^b	1.44	0.87; 0.83	
Gly 10	8.40	—	4.02; 3.84				

^a(S) and (R) refer to the pro-S or R Leu- C^βH_2 protons assigned from the NOEs in DMSO (16); (c) and (t) refer to the *cis* or *trans* disposition of the Pro- C^βH_2 protons with respect to the carbonyl.

^bDue to the “free” rotation of the Leu 9 C^α - C^β bond, the pro-R or S Leu 9- C^βH_2 protons are not assigned.

^{13}C resonances: The assignment of the ^{13}C resonances can give additional conformational information. For example, in the absence of the NH amide proton in N-methylamino acid residues and therefore of the ^1H , ^1H vicinal coupling constant, the vicinal heteronuclear coupling constants between the N^{13}CH_3 carbon and the α -proton provides an estimation of the ϕ dihedral angle (17). The *cis* or *trans* conformation of the Xaa-Pro peptide bond is also directly related to the ^{13}C chemical shifts for the proline ring (18). Moreover, the $^{13}\text{C}/^1\text{H}$ heteronuclear correlation via the two- and three-bond scalar couplings involving the ^{13}CO carbon gives access to the sequential assignment.

The aliphatic ^{13}C resonances have been assigned by heteronuclear shift correlated 2D-experiments (19) using the selective polarization transfer from ^1H to ^{13}C -nuclei via the $^1J(^{13}\text{C}-^1\text{H})$ direct coupling with an average value of 135 Hz. The Pro- $^{13}\text{C}/^1\text{H}$ connectivities are indicated as an example in figure 4. The presence of the *cis* Val 7-Pro 8 peptide bond is confirmed by the clear non-equivalence by 9.5 ppm of the Pro- $^{13}\text{C}^\beta$ and $^{13}\text{C}^\gamma$ resonances. A *trans* peptide bond would correspond to a difference smaller than 5 ppm (18).

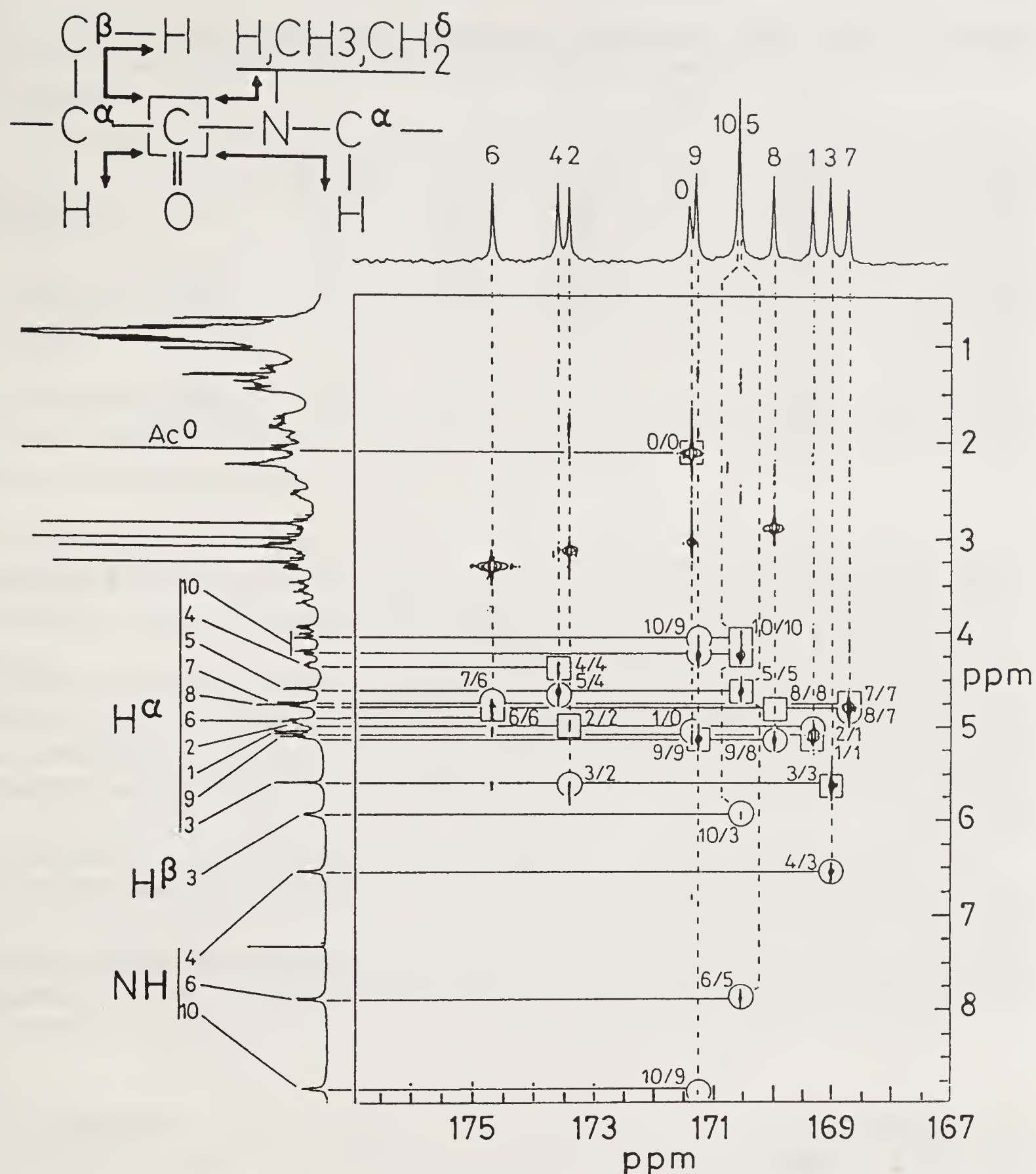


Figure 5. $^1\text{H},^{13}\text{C}$ -COSY spectrum for the carbonyl region of griselimycin in chloroform. The complete sequence is given by the $^1\text{H}/^{13}\text{C}$ connectivities.

when the corresponding $\Delta\delta/\Delta T$ value is below 3×10^{-3} ppm/K. In CDCl_3 solution (figure 6), the small temperature gradients for all three NH protons are typical of hydrogen-bonded protons. However, the small chemical shift of the Leu 4-NH proton indicates a weaker interaction than for the two other cases. In DMSO solution (figure 6), the temperature coefficients for both Leu 6 and Gly 10-NH resonances remain small, but the high $\Delta\delta/\Delta T$ value for the Leu 4-NH resonance is indicative of a solvent-exposed N-H bond.

In order to investigate the solvent effect on the NH proton signals, DMSO has been progressively added to a CDCl_3 solution while the peptide concentration was kept constant. It can be assumed that a hydrogen-bonded amide proton is less readily solvated by DMSO than a free amide proton and its signal is accordingly less rapidly

Table 2. ¹³C-NMR chemical shifts (in ppm from TMS) of griselimycin in CDCl₃.

	CO	NCH ₃	C ^α	C ^β	C ^γ	C ^δ	Other
CH ₃ CO	171.4						
NMe-L-Val 1	169.3	31.5	58.9	27.7	18.9; 18.8		
Me _{4t} -L-Pro 2	173.4		56.5	36.4	32.2	54.4	17.6 (Me _{4t})
NMe-L-Thr 3	169.0	34.2	59.9	69.2	17.8		
L-Leu 4	173.6		51.0	40.0	25.2	23.4; 20.8	
Me _{4t} -L-Pro 5	170.5		60.3	33.1	33.6	53.3	16.7 (Me _{4t})
L-Leu 6	174.7		46.5	41.3	24.6	23.5; 21.5	
NMe-L-Val 7	168.7	30.7	57.7	20.3	19.0; 18.3		
L-Pro 8	170.0		58.6	31.5	22.0	46.9	
NMe-D-Leu 9	171.3	30.6	55.3	37.0	24.7	22.9; 22.6	
Gly 10	170.5		42.8				

shifted. Among the three NH amide proton in griselimycin, both the Leu 6 and Gly 10-NH signals are shifted by less than 0.5 ppm while a large shift (about 1.3 ppm) is observed for the Leu 4-NH resonance (figure 6).

These observations are corroborated by the N-H stretching absorption in CHCl₃ which shows an intense component at 3287 cm⁻¹ typical of the strongly hydrogen-bonded Leu 6 and Gly 10-NH bonds, and a smaller contribution at 3371 cm⁻¹ which can be assigned to the Leu 4-NH bond engaged in a much weaker interaction.

When considering the carbonyl ¹³C signals, the Leu 6, Leu 4 and Pro 2 resonances are particularly deshielded (figure 5), which suggests their hydrogen-bonded character (26).

The Leu 6-CO seems to be the most adequate accepting site from the Gly 10-NH bond. It allows the close *trans*-annular contacts mentioned above, and the resulting

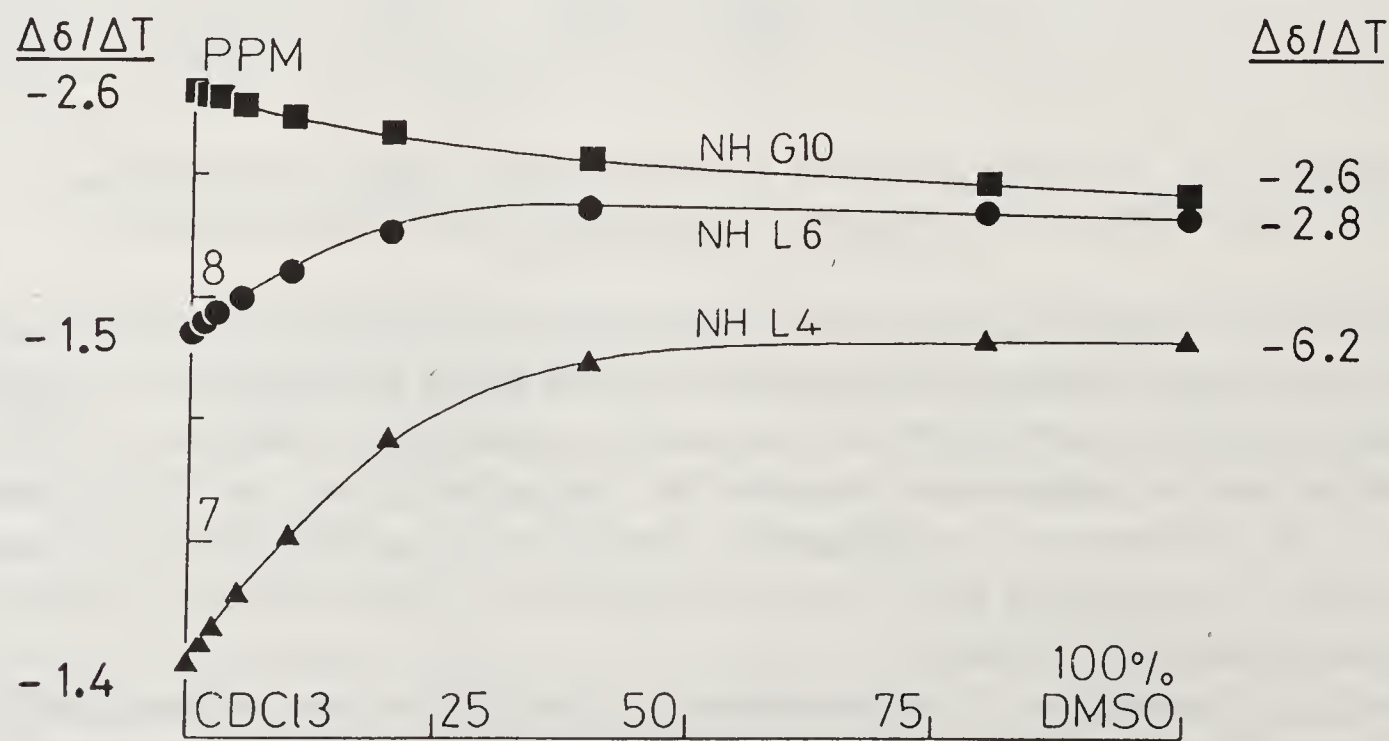


Figure 6. Solvent dependence of the NH proton NMR signals of griselimycin as a function of DMSO content in DMSO-*d*₆/CDCl₃ mixtures. The temperature dependence $\Delta\delta/\Delta T$ is also indicated in 10⁻³ ppm/°C for both pure solvents.

C₁₃ folded structure encompassing the *cis* Val 7–Pro 8 amide bond has been already pointed out in other bioactive cyclopeptides such as antanamide (27) and cyclinopeptide A (28).

We suggest that the Leu 6-NH is engaged in a bifurcated interaction with the Leu 6 and Leu 4 carbonyls, resulting in a so-called C₇C₅ conformation already found in the Pro–Ala sequence (29), and compatible with the Leu 6-NH stretching frequency at 3287 cm^{−1}.

The Leu 4-NH could be engaged in a weaker interaction with the Pro 2 carbonyl, and the flexibility of the acyclic chain could account for its solvent sensitivity to DMSO.

Conformation of the N-C^α bonds: The vicinal H-N-C^α-H coupling constant is related to the dihedral angle θ by the Karplus equation (30):

$$^3J(\text{HN-C}^\alpha\text{H}) = A \cos 2\theta - B \cos \theta + C$$

in which A , B and C depend on the system considered. θ is related to the rotational angle ϕ (31) by:

$$\theta = |\phi - 60^\circ| \quad \text{for L residues}$$

and

$$\theta = |\phi + 60^\circ| \quad \text{for D residues.}$$

An empirical calibration of the A , B and C coefficients for the treatment of peptides has been undertaken simultaneously by Bystrov *et al.* (32, 33), by Ramachandran *et al.* (34) and by ourselves (35, 36). In the following, we have used our own calibration with $A = C = 4.3$ and $B = 2.9$ Hz.

A great number of antibiotic peptides extracted from plants or micro-organisms (2–13, 37) contain non-coded N-methylamino-acid residues so that the commonly used $^3J(\text{HN-C}^\alpha\text{H})$ coupling is not available. For these residues, a previous study has shown that the $^3J(^{13}\text{CH}_3\text{N-C}^\alpha\text{H})$ heteronuclear vicinal coupling constant is also related by a Karplus relation to the same dihedral angle θ as above by (38):

$$^3J(^{13}\text{CH}_3\text{N-C}^\alpha\text{H}) = 3.0_5 \cos 2\theta - 2.5_5 \cos \theta + 3.2$$

The $^3J(\text{HN-C}^\alpha\text{H})$ and $^3J(^{13}\text{CH}_3\text{N-C}^\alpha\text{H})$ coupling constants and the resulting ϕ angles are given in table 3. Except for the NMe-D-Leu 9 ($\phi = 75^\circ$ or 165°) and Gly 10 ($\phi = \pm 60^\circ$) residues, all the ϕ angles are equal to $-80^\circ \pm 10^\circ$ or $-160^\circ \pm 10^\circ$. One notes some small changes from CDCl₃ to DMSO, when the weak intramolecular interaction involving the Leu 4-NH amide hydrogen is broken.

Conformation of the side-chains: Usually, in the C^αH-C^βH₂ fragment as in Leu residues, the two vicinal coupling constants $^3J_{\alpha\beta\text{R}}$ and $^3J_{\alpha\beta\text{S}}$ are interpreted in terms of three staggered rotamers (39). The populations $p\text{I}$, $p\text{II}$ and $p\text{III}$ of rotamers I ($\chi^1 = -60^\circ$), II ($\chi^1 = 180^\circ$) and III ($\chi^1 = 60^\circ$) are expressed by:

$$p\text{I} = (^3J_{\alpha\beta\text{R}} - ^3J_g) / (^3J_t - ^3J_g)$$

$$p\text{II} = (^3J_{\alpha\beta\text{S}} - ^3J_g) / (^3J_t - ^3J_g)$$

$$p\text{III} = (^3J_t + ^3J_g - ^3J_{\alpha\beta\text{R}} - ^3J_{\alpha\beta\text{S}}) / (^3J_t - ^3J_g)$$

The 3J_t and 3J_g coupling constants correspond to the *trans* and *gauche* dispositions of the H^α and H^β protons, and have been taken equal to 13.9 and 3.55 Hz, respectively

Table 3. Estimation of the ϕ dihedral angles in CHCl_3 and DMSO on the basis of Karplus relationships. Comparison with the conformational angles deduced from the theoretical analysis.

Residue	Fragment	3J (Hz)		$\phi(^{\circ})$		Theoretical analysis	
		CDCl_3	DMSO	CDCl_3	DMSO	$\phi(^{\circ})$	$\psi(^{\circ})$
NMe-L-Val 1	$^{13}\text{CH}_3\text{NC}^{\alpha}\text{H}^{\text{a}}$	4.4	c	−70, −170	c	−120	60
Me _{4t} -L-Pro 2						−69	168
NMe-L-Thr 3	$^{13}\text{CH}_3\text{NC}^{\alpha}\text{H}^{\text{a}}$	5.9	c	−80, −160	c	−130	−162
L-Leu 4	$\text{HNC}^{\alpha}\text{H}^{\text{b}}$	7.7	5.3	−80, −160	−70, −170	−66	139
Me _{4t} -L-Pro 5						−65	84
L-Leu 6	$\text{HNC}^{\alpha}\text{H}^{\text{b}}$	9.4	7.4	−90, −150	−75, −165	−158	89
NMe-L-Val 7	$^{13}\text{CH}_3\text{NC}^{\alpha}\text{H}^{\text{a}}$	4.5	c	−70, −170	c	−100	142
L-Pro 8						−69	155
NMe-D-Leu 9	$^{13}\text{CH}_3\text{NC}^{\alpha}\text{H}^{\text{a}}$	4.7	c	75, 165	c	133	−86
Gly 10	$\text{HNC}^{\alpha}\text{H}_2^{\text{b}}$	3.9; 6.4	4.2; 6.3	± 60	± 60	72	−140

^aSee reference (38) ^bSee reference (36) ^cnot determined.

(39). However, the pro-*R* or *S* attribution of the $\text{C}^{\beta}\text{H}_2$ protons is often questionable, so that the rotamers I and II cannot be easily discriminated. To this aim, we have combined the $^3J_{\alpha\beta}$ coupling constants and the intensity of the $\text{NH}/\text{C}^{\beta}\text{H}$ NOESY intraresidue correlations. Rotamer I only gives one short $\text{NH}/\text{C}^{\beta}\text{H}_R$ interproton contact whereas rotamer II is characterized by two short $\text{NH}/\text{C}^{\beta}\text{H}_R\text{H}_S$ interproton contacts (16).

In proline residues, the five-membered ring can adopt a continuous range of conformational states between two extreme conformations characterized by $\chi^1 \sim \pm 40^{\circ}$ (40, 41). The rigid or fluctuating character of the ring can be deduced from the $^3J_{\alpha\beta}$ coupling constants (41, 42).

The Val residue has a single C^{β}H proton and a single $^3J_{\alpha\beta}$ coupling constant so that only a global estimation of the rotamers II and III is possible.

The $^3J_{\alpha\beta}$ coupling constants obtained from the *J*-resolved spectrum are listed in table 4. We note first that, excepting the Pro 2 and Leu 9 residues, the side-chains are similarly oriented in CHCl_3 and DMSO.

The five-membered ring in both Pro 5 and Pro 8 is rigid with $\chi^1 \sim 30^{\circ}$. The same holds true for the Pro 2 residue in CDCl_3 , but it becomes flexible in DMSO. Excepting the D-Leu 9 side-chain, all the side-chains assume a preferential orientation probably due to steric hindrances. Thus the rotamer I ($\chi^1 = -60^{\circ}$) is largely predominant for both Val 1 and Val 7 residues. Similarly, the Leu 4 and Leu 6 side-chains also assume the preferential rotamer I and II, respectively. The small $^3J_{\alpha\beta}$ coupling constant for the L-Thr 3 side-chain probably corresponds to the rotamer III which is by far the most frequent one in the crystallized peptides and proteins (43, 44).

Nuclear Overhauser enhancements: Cyclic peptides are convenient compounds to observe short interproton distances which mostly arise from intraresidue or nearest-neighbour residue connectivities. Usually, the NOEs give structural information in terms of the short-range distances $d_{\text{N}\alpha}$ (from NH_i to $\text{C}^{\alpha}\text{H}_i$), $d_{\alpha\text{N}}$ (from NH_i to $\text{C}^{\alpha}\text{H}_{i-1}$), d_{NN} (from NH_i to either NH_{i+1} or NH_{i-1}) and $d_{\text{N}\beta}$ (from NH_i to

Table 4. Rotational states of the side-chains in chloroform and DMSO.^a

Residue	Protons ^b	CDCl ₃		DMSO	
		³ J _{αβ}	Rotational state (%)	³ J _{αβ}	Rotational states (%)
NMe-L-Val 1	αβ	11.1	pI = 75 pII + pIII = 25	10.9	pI = 70 pII + pIII = 30
Me _{4t} -L-Pro 2	αβ _c	2.4	rigid cycle	4.4	rapid interconversion χ ¹ ~ ± 30°
	αβ _t	8.8	χ ¹ ~ 30°	7.5	
NMe-L-Thr 3	αβ	3.2	pIII = 100	3.5	pIII = 100
L-Leu 4	αβ _S	≤ 1.5	pI = 100	≤ 1.5	pI = 100
	αβ _R	11.7	pII = pIII = 0	10.4	pII = pIII = 0
Me _{4t} -L-Pro 5	αβ _c	≤ 1.5	rigid cycle	≤ 1.5	rigid cycle
	αβ _t	8.0	χ ¹ ~ 30°	8.5	χ ¹ ~ 30°
L-Leu 6	αβ _S	10.2	pI = 5	10.4	pI = 0
	αβ _R	4.3	pII = 65 pIII = 30	3.7	pII = 70 pIII = 30
NMe-L-Val 7	αβ	10.8	pI = 70	10.6	pI = 70
			pII + pIII = 30		pII + pIII = 30
L-Pro 8	αβ _c	2.0	rigid cycle	≤ 1.5	rigid cycle
	αβ _t	8.5	χ ¹ ~ 30°	8.3	χ ¹ ~ 30°
NMe-D-Leu 9	αβ ^c	7.2	pI = pII = 35	5.4	pI (or pII) = 50
	αβ ^c	7.2	pIII = 30	9.5	pII (or pI) = 35 pIII = 25

^aRotamer I: χ¹ = -60°; II: χ¹ = 180°; III: χ¹ = 60°.

^b(S) and (R) refer to the pro-S or R Leu-C^βH₂ protons assigned from the NOEs in DMSO (16); (c) and (t) refer to the *cis* or *trans* disposition of the Pro-C^βH₂ protons with respect to the carbonyl.

^cDue to the "free" rotation of the Leu 9 C^α-C^β bond, the pro-R or S Leu 9-C^βH₂ protons are not assigned.

C^βH_{i-1}). Some of these distances are directly related to the usual torsional angles, namely *d*_{Nα} to φ, *d*_{αN} to ψ, and *d*_{NN} to both φ and ψ (45, 46).

Due to the presence of various proline and N-methylamino acid residues, no *d*_{NN} connectivity is observed in griselimycin. However, some correlations involving the NMe protons are found as for the Leu 9-NMe and Gly 10-NH protons for example.

NOESY experiments in DMSO, in phase sensitive mode with seven different mixing times varying from 100 to 600 ms, have been performed in order to check the linearity of the NOE build up. The cross-relaxation rates are maximum for a mixing time of about 500 and 600 ms. The great number of strong short- and medium-range NOE cross-peaks between the backbone-backbone, backbone-side chain and side chain-side chain proton signals argue in favor of a rigid folded structure. The NOEs are negative, suggesting that the rotational correlation time τ_c is long with ωτ_c > 1 (47). All the NOEs have been quantified from their linear variation with the mixing time, and analyzed in terms of interproton distances taking the Gly 10-C^αH₂ interproton distance as a reference.

An important point is the close intra-annular proximity of both Gly 10 α-protons with the Pro 5 α, Leu 6-NH and Val 7 α-protons. A weaker cross-peak is also observed

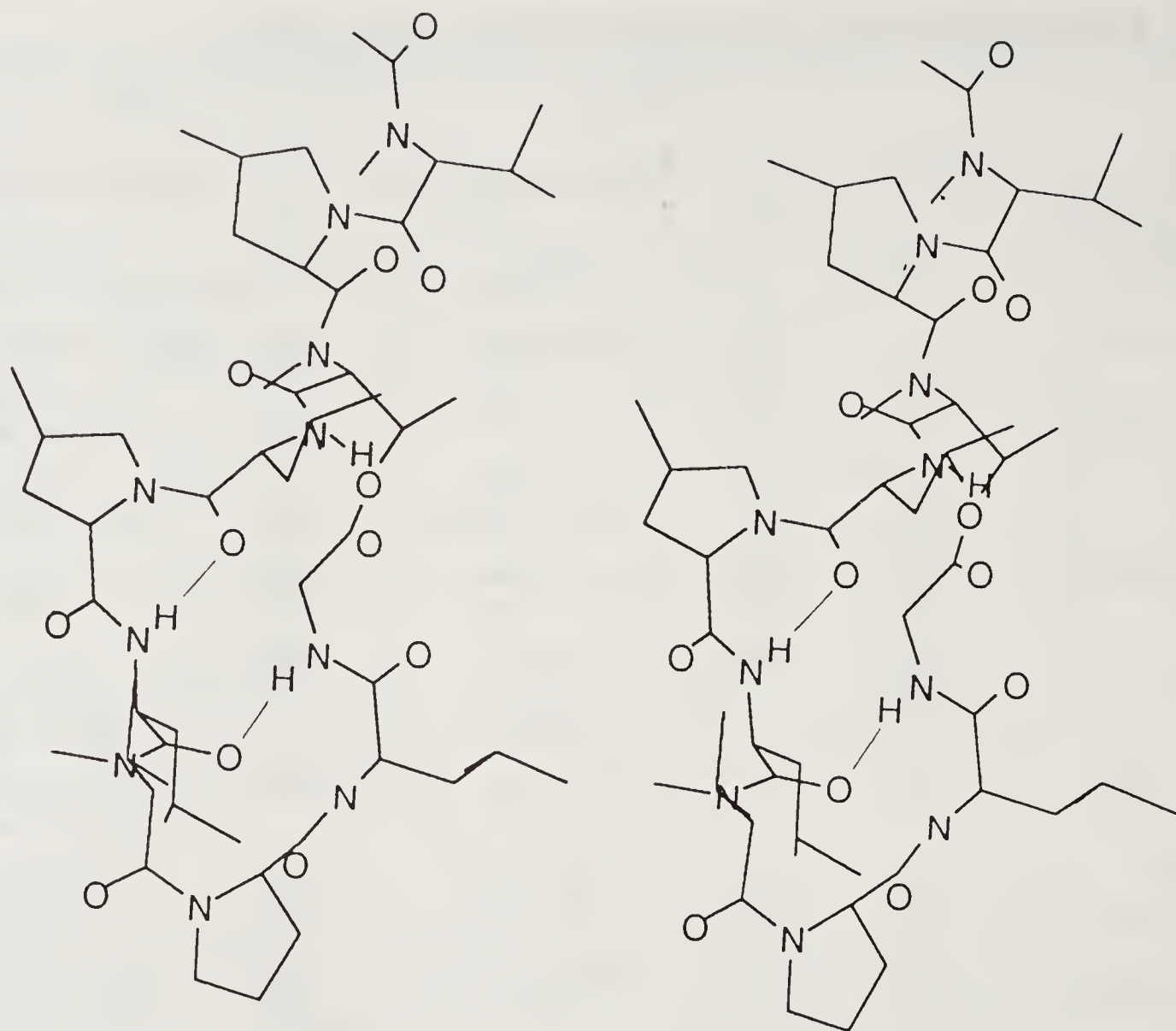


Figure 7. Stereoviews of the conformation of griselimycin in dimethylsulfoxide given by molecular modeling of the conformational constraints determined by 2D-NMR experiments. The two intra-annular hydrogen bonds are indicated in thin lines.

between the Gly 10-NH and Leu 6- $C^{\beta}H_2$ protons. These data indicate a folded structure for the NMe-L-Val 7-L-Pro 8-NMe-D-Leu 9 sequence.

Molecular dynamics: The conformation of griselimycin has been studied by molecular dynamics (MD) simulation to model the structure of this compound in DMSO solution. The energy minimization and retrained MD have been performed using a modified version of AMBER 3.0 program (48), and SYBYL (49) and ADAPTU (50) programs for molecular display.

Griselimycin has been built by minimization and dynamic simulation of the linear conformation using the NOEs as distance constraints, and a ring closing bond between Gly 10-CO and Thr 3-OH fixed at 1.425 Å. In the case of the NOEs involving N-methyl groups, the distance constraint has been placed on the N-methyl carbon while increasing by 0.3 Å the uncertainty on the interproton distance.

The minimized geometry has been used as starting coordinates and MD has been started by taking the initial velocities for the atoms from a Maxwellian distribution at 300 K. Non bonded interactions have been evaluated within a 10 Å cutoff with updating the pair list every 50 steps and choosing a dielectric constant equal to 50. The first 5ps of the MD simulation have been used to equilibrate the system. After this time, the temperature has been gradually increased (100 K/1ps) up to 1000 K.

MD simulation at 1000 K has been performed for 80 ps in order to increase the ability to cross higher energy barriers and generate conformers. Details will be presented elsewhere.

Several conformers have been selected on the basis of the NOE data concerning the peptide backbone. After energy minimization, one of these conformers (figure 7) is mostly in agreement with the backbone dihedral angles ϕ deduced from the NH-C α H coupling constants (table 4). The structure of griselimycin proposed shows three intra-annular hydrogen bonds. The first interaction involving Gly 10-NH and Leu 6-CO closes a 13-membered cycle similar to that inferred from the NMR data. The second one concerns Leu 6-NH and Leu 4-CO and closes a 7-membered cycle around Me₄-Pro 5, a residue which has probably the same γ -turn inducing properties as proline. The Leu 6- ψ value of about 90° is not compatible with the fused C₇C₅ structure proposed on the basis of the IR and NMR data in CHCl₃. In the present model, the Leu 4-NH bond is free of intramolecular contact, in good agreement with the NMR data collected in DMSO.

Conclusion

The conformational analysis of griselimycin reported here is based on the unambiguous assignment of the ¹H and ¹³C resonances and the IR data.

In spite of the presence of numerous tertiary amide bonds, a single conformational state with a *cis* NMe-Val 7-Pro 8 amide bond is observed. In chloroform, griselimycin adopts a rigid conformation stabilized by three hydrogen bonds of the N-H...O=C type with formation of a C₇C₅ structure (inversed γ -turn) and a C₁₃ structure (related to the well-known α -turn). The Leu 6 carbonyl which participates in the *trans*-annular hydrogen bond is probably the common accepting site from the Leu 6-NH and Gly 10-NH bonds. The Leu 6-NH is also involved in a bifurcated hydrogen bond with the Leu 4 and Leu 6 carbonyls. These intra-annular contacts, with the probable exception of the C₅ conformation of the Leu 6 residue, are retained in DMSO. The third hydrogen bond in chloroform concerns the Leu 4-NH and Pro 2-CO and also closes a 7-membered cycle. Due to the location of Pro 2 in the extracyclic, and therefore rather flexible, part of the molecule, this interaction is weaker than the two others and is not retained in a strong solvating medium such as DMSO.

This structure is compatible with the NOE data as distance constraints in a molecular dynamic simulation using AMBER program. Work is in progress in order to consider more precisely the orientation of the side-chains and the flexibility of the proline 5-membered ring.

The most salient fact of this study is the C₁₃ structure involving the Leu 6-NMe-Val 7-Pro 8-NMe-D-Leu 9-Gly 10 fragment with the *trans-cis-trans-trans* amide bond sequence. It seems to be particularly favoured in large cyclopeptides as it has been already characterized in the crystal structures of antanamide and cyclinopeptide A.

Acknowledgements

We are indebted to the Rhone-Poulenc Society for the generous gift of griselimycin. We are also very grateful to Drs. Folkers and Ptak for their interest in this work.

References

1. Terlain, B. & Thomas, J.-P. (1971) *Bull. Soc. Chim. France* 2349–2356, 2357–2362, 2363–2365.
2. Jolad, S. D., Hoffmann, J. J., Torrance, S. J., Wiedhopf, R. M., Cole, J. R., Arora, S. K., Bates, R. B., Gargiulo, R. L. & Kriek, G. R. (1977) *J. Am. Chem. Soc.* **99**, 8040–8044.
3. Meyer, W. L., Templeton, G. E., Grable, C. I., Jones, R., Kuyper, L. F., Lewis, R. B., Sigel, C. W. & Woodhead, S. H. (1975) *J. Am. Chem. Soc.* **97**, 3802–3809.
4. Traber, R., Loosli, H. R., Hofmann, H., Kuhn, M. & Von Wartburg, A. (1982) *Helv. Chim. Acta* **65**, 1655–1677.
5. Ovchinnikov, Yu. A., Ivanov, V. T., Evstratov, A. V., Mikhaleva, I. I., Bystrov, V. F., Portnova, S. L., Balashova, T. A., Meshcheryakova, E. N. & Tulchinsky, V. M. (1974) *Int. J. Pept. Protein Res.* **6**, 465–498.
6. Jain, S. C. & Sobell, H. M. (1972) *J. Mol. Biol.* **68**, 1–20.
- 7–Pro 8-NMe-D-Leu 9–Gly 10 fragment with the *trans*–*cis*–*trans*–*trans* amide bond
8. Vanderhaeghe, H. & Parmentier, G. (1960) *J. Am. Chem. Soc.* **82**, 4414–4422.
9. Anteunis, M. J. O., Callens, R. E. A. & Tavernier, D. K. (1975) *Eur. J. Biochem.* **58**, 259–268.
10. Sheehan, J. C., Zachau, H. G. & Lawson, W. B. (1957) *J. Am. Chem. Soc.* **79**, 3933–3934.
11. Bodanszky, M., Izdebski, J. & Muramatsu, I. (1969) *J. Am. Chem. Soc.* **91**, 2351–2358.
12. Hassall, C. H., Thomas, W. A. & Moschidis, M. C. (1977) *J. Chem. Soc., Perkin Trans. I*, 2369–2376.
13. Kessler, H., Mronga, S., Will, M. & Schmidt, U. (1990) *Helv. Chim. Acta* **73**, 25–47.
14. Stverteczky, J. & Bajusz, S. (1976) *Acta Chim. Acad. Sci. Hung.* **88**, 67–74.
15. Wider, G., Macura, S., Kumar, A., Ernst, R. R. & Wüthrich, K. (1984) *J. Magn. Reson.* **56**, 207–234.
16. Kessler, H., Griesinger, C. & Wagner, K. (1987) *J. Am. Chem. Soc.* **109**, 6927–6933.
17. Cung, M. T., Boussard, G., Vitoux, B. & Marraud, M. (1980) *C. R. Acad. Sci.* **C290**, 291–294.
18. Siemion, I. Z., Wieland, T. & Pook, K.-H. (1975) *Angew. Chem. Int. Ed. Engl.* **14**, 702–703.
19. Bax, A. & Morris, G. A. (1981) *J. Magn. Reson.* **42**, 501–505.
20. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963) *J. Mol. Biol.* **7**, 95–99.
21. Ramachandran, G. N., Ramakrishnan, C. & Venkatachalam, C. (1965) *Biopolymers* **3**, 591–592.
22. Ramakrishnan, C. & Ramachandran, G. N. (1965) *Biophys. J.* **5**, 909–933.
23. Ramachandran, G. N., Venkatachalam, C. & Krimm, S. (1966) *Biophys. J.* **6**, 849–872.
24. Ohnishi, M. & Urry, D. W. (1969) *Biochem. Biophys. Res. Commun.* **36**, 194–202.
25. Llinas, M., Klein, M. P. & Neilands, J. B. (1970) *J. Mol. Biol.* **52**, 399–414.
26. Loosli, H. R., Kessler, H., Oschikmat, H., Weber, H. P., Petcher, T. J. & Widmer, A. (1985) *Helv. Chim. Acta* **68**, 682–704.
27. Karle, I. L., Wieland, T., Schermer, D. & Ottenheim, H. C. J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1532–1536.
28. Di Blasio, B., Rossi, F., Benedetti, E., Pavone, V., Pedone, C., Temussi, P.A., Zanotti, G. & Tancredi, T. (1989) *J. Am. Chem. Soc.* **111**, 9089–9098.
29. Boussard, G., Cung, M. T., Marraud, M. & Néel, J. (1974) *J. Chim. Phys., Phys.-Chim. Biol.* **71**, 1159–1166.
30. Karplus, N. M. (1959) *J. Chem. Phys.* **30**, 11–15.
31. IUPAC-IUB Commission on Biochemical Nomenclature (1970) *Biochemistry* **9**, 3471–3479.
32. Bystrov, V. F., Portnova, S. L., Tsetlin, V. T., Ivanov, V. T. & Ovchinnikov, Yu. A. (1969) *Tetrahedron* **25**, 493–515.
33. Bystrov, V. F., Ivanov, V. T., Portnova, S. L., Balashova, T. A. & Ovchinnikov, Yu. A. (1973) *Tetrahedron* **29**, 873–877.
34. Ramachandran, G. N., Chandrasekaran, R. & Kopple, K. D. (1971) *Biopolymers* **10**, 2113–2131.
35. Cung, M. T., Canet, D., Granger, P., Marraud, M. & Néel, J. (1969) *C.R. Acad. Sci.* **C269**, 580–583.
36. Cung, M. T., Marraud, M. & Néel, J. (1974) *Macromolecules* **7**, 606–613.
37. Kalman, J. R., Blake, T. J., Williams, D. H., Feeney, J. & Roberts, G. C. K. (1979) *J. Chem. Soc., Perkin Trans. I*, 1313–1321.

38. Cung, M. T., Boussard, G., Vitoux, B. & Marraud, M. (1980) *C.R. Acad. Sci.* **C290**, 291–294.
39. Cung, M. T. & Marraud, M. (1982) *Biopolymers* **21**, 953–967.
40. Nair, C. M. K. & Vijayan, M. (1981) *J. Indian Inst. Sci.* **C63**, 81–103.
41. Cung, M. T., Vitoux, B. & Marraud, M. (1987) *New J. Chem.* **6**, 503–510.
42. Bach II, A. C., Bothner-By, A. A. & Gierasch, L. M. (1982) *J. Am. Chem. Soc.* **104**, 572–576.
43. Janin, J., Wodak, S., Lévitte, M. & Maigret, B. (1978) *J. Mol. Biol.* **125**, 357–386.
44. Benedetti, E., Morelli, G., Nemethy, G. & Scheraga, H. A. (1983) *Int. J. Pept. Protein Res.* **22**, 1–15.
45. Billeter, M., Braun, W. & Wüthrich, K. (1982) *J. Mol. Biol.* **155**, 321–346.
46. Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids* (John Wiley & Sons, New York).
47. Balaram, P., Bothner-By, A. A. & Dadok, J. (1972) *J. Am. Chem. Soc.* **95**, 4015–4017.
48. Weiner, P. & Kollman, P. A. (1981) *J. Comput. Chem.* **2**, 287–303.
49. TRIPOS Associates Inc., Saint Louis, USA.
50. Krug, M. & Folkers, G.: Animated Analysis Program at Tübingen University, Pharmaceutical Institute, Tübingen, FRG.

Structure and Function of Transmembrane Ion Channels

Robert M. Stroud[†]

Department of Biochemistry and Biophysics, University of California, San Francisco,
CA 94143, USA

Introduction

G. N. Ramachandran's contributions in the development of a quantitative method for looking at peptide conformations are only now beginning to be truly appreciated, and used to greater effect. It has become a revelation to discover the wide ranging applicability of these ideas. The slowness of our appreciation should not hide the revolutionary nature of his insight, far ahead of its time, that gave rise to these ideas. Only now, in 1990, with the increasing number of highly refined protein structures determined by X-ray crystallography, is it realized that the conformations of peptides within a protein chain follow those predicted by Ramachandran with an agreement that increases with every refinement cycle. Now, in 1990, these observations of the macromolecular crystallographers with large protein molecules are being recognized as the basis for both the assessment of the state of refinement of a protein structure, and as a possible constraint that could be applied in the determination of protein structures. The 'Ramachandran plot' of angles ϕ versus ψ has truly become accepted as the standard for judgement, and for expectation. Shown in figure 1 is the Ramachandran plot for trypsin, one of the most highly refined protein structures, currently refined in my laboratory by the difference Fourier method to a residual of 13%, at a resolution of 1.3 Å (1). The 220 peptide orientations fall beautifully within the values predicted by Ramachandran, with no bias having been introduced in this direction during refinement. Perhaps macromolecular crystallographers did not expect large proteins to conform to the energetics of peptides, with their much larger energy of folding. But they do! And now this information can be used in the rapid assessment of protein structures, prediction of new structures and in the determination of protein structures as well.

The value in prediction of protein structures is seen particularly in membrane proteins, whose structures have been unusually difficult to determine due to difficult crystallizations. Here the need to satisfy electrostatic forces including hydrogen bonds within the membrane is high, and the regular helix and sheet structures are forced more strongly upon the polypeptide chains within the hydrophobic milieu. Thus

[†]Supported by the National Science Foundation PCM83-16401, and National Institutes of Health GM24485 to RMS.

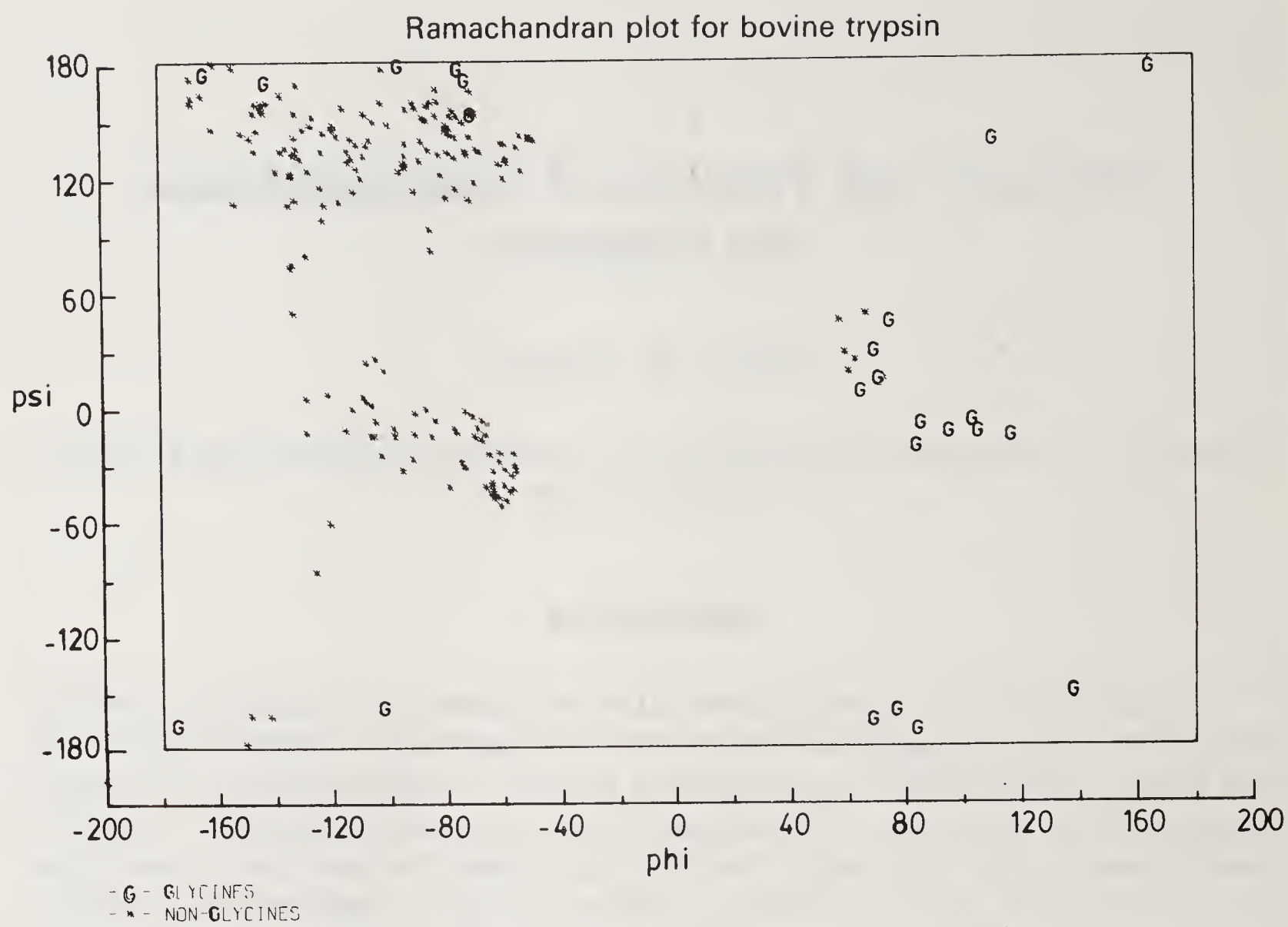


Figure 1. Ramachandran plot showing the ϕ , ψ angles observed for the 220 amino acids in the highly refined structure of bovine trypsin, determined in the authors laboratory and refined to a residual of 13% at 1.3 Å resolution.

within the membrane, perhaps more than anywhere else, there is both the energetic pressure to conform, and the need for any viable means of making predictions of conformation.

Ligand-gated ion channels provide for the rapid dialogue between cells of the central nervous system, converting a chemical neurotransmitter signal released from one cell into an electrical signal that propagates along the target cell membrane. To date, three types of ligand-gated receptors that mediate information transfer in the brain and neuromuscular junction have been characterized at the level of their sequences. They are the excitatory nicotinic AChRs of neuromuscular and neuronal origins which conduct cations, excitatory neuronal kainate type glutamate-activated channels which conduct cations, and the inhibitory neuronal GABA and glycine-activated channels which conduct chloride. Different in size and sequence, these types of channels comprise a super-family whose subunits share several features: four putative membrane spanning regions, known to form α -helices, denoted M1-M4; strong sequence homology in these hydrophobic transmembrane sequences, especially in M2, the second hydrophobic stretch of ~ 20 amino acids that probably lines the ion-conducting channel; and probable quasi-symmetric pentameric arrangement of similar subunits. In addition, a homologous cystine-bridged loop is found in these cell surface receptor subunits. Also common among these signalling families of the brain or neuromuscular junction is the way diversity within a family seems to be provided for by expression of genes that substitute for one or other of the subunits

that together form functional receptors. How do these ligand-gated ion channels function? Already, in the absence of high-resolution structures, much has been deduced about how their structure supports function.

The Nicotinic AChR Family

Nicotinic AChRs conduct $\sim 10^4$ sodium ions per ms in response to binding of cholinergic agonists. Channels opening follows less than $10\ \mu\text{s}$ after ligand binding and the flow sodium ions into the cell rapidly short-circuits the normal ionic gradient maintained across the plasma membrane by ATP driven pumps, and so depolarizes it. The typically $\sim -50\ \text{mV}$ potential difference across the membrane is reduced to $\sim -5\ \text{mV}$, thus transducing a chemical signal into an electrical one that signals muscle contraction at the neuromuscular junction. Individual channel openings are brief, averaging 1 ms for junctional receptors at a holding potential of $-100\ \text{mV}$ in the frog neuromuscular junction.

Electrocytes which compose the electric organ of the electric rays *Torpedo* and *Narcine*, and electric eel *Electrophorus*, contain close-packed AChR in post-synaptic regions of the plasma membrane. Because AChR can be isolated in large quantity from electric organs it has yielded most of the biochemical insight into the AChR family (2, 3, 25).

A Pentamer of Homologous Subunits Surrounds a Channel

The AChR is a complex of five homologous transmembrane glycoprotein subunits. There are four gene products designated α , β , γ , and δ molecular weights 52.4, 56.2, 63.2, 65.9 kDa. There is high sequence homology (35–50%), and therefore evolutionary relatedness, among the four subunit types (3). In 1982 four groups reported cloning and partial or complete sequencing of cDNAs of individual AChR subunits from either *T. californica* or *T. marmorata* (4–7). After alignment, sequences of α (4, 5, 8) β , δ , (9) and γ subunits (10, 11) show 19% identity and 54% homology (10).

AChR from *Torpedo* forms ordered two-dimensional arrays (12, 13), and ordered tubular vesicles with crystalline lattices. These were used for two-dimensional image reconstruction (14) and combined with X-ray diffraction to give a composite three-dimensional structure (15). AChR in these tubular membrane vesicles crystallizes with dimeric symmetry (16, 17) which is independent of the covalent bonds formed 80% of the time (17) between C-termini of the δ chains of adjacent molecules (18). These crystals gave rise to three-dimensional structures from tilted samples at $25\ \text{\AA}$ both in stain and in ice (19) from helical diffraction (20) and by combined tilting stage electron microscopy at $22\ \text{\AA}$ and X-ray diffraction to $12.5\ \text{\AA}$ resolution perpendicular to the membrane plane (17) (figures 2, 3).

Two Different Binding Sites Optimize Timing

The α chains lie in non-equivalent environments, separated by an angle of $144^\circ \pm 4^\circ$, precisely two sectors of a pentagon, around the axis (23); they are not adjacent in

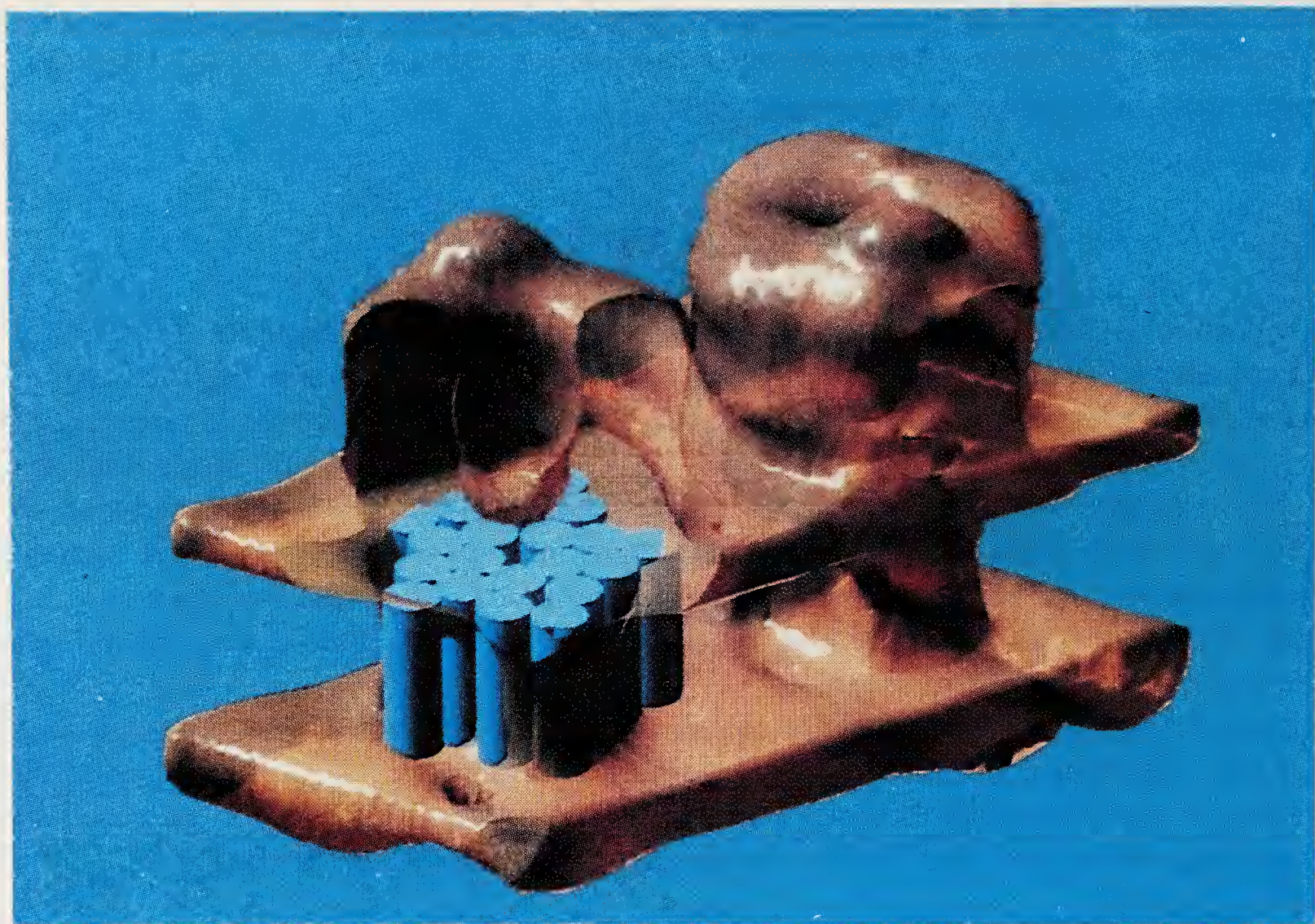


Figure 2. Outer surface of the AChR structure seen from a 3-dimensional electron microscopic and X-ray diffraction structure analysis to $22 \text{ \AA} \times 12.5 \text{ \AA}$ resolution (17). The surfaces enclose positive densities. The density for the bilayer, in blue, appears as horizontal planes of densities which correspond to the dense phosphatidyl head groups that are 41 \AA apart (21). The bundles of cylinders, shown to scale, have the dimensions of 30 \AA length close-packed α -helices in the hydrophobic center of the bilayer-spanning region, as suggested by X-ray diffraction and sequence analysis, equivalent to ~ 20 amino acid residues in length, and are shown in a cutaway for one of the AChR molecules in the dimer. Relative to the entry well the closest helices are most likely M2, left handed supertwisted, based on model building (22). The synaptic side of these helices form the base of the well. The 25 \AA diameter infundibulum extends from the outermost extremity to just below the surface level of the phospholipid bilayer.

the pentamer. Binding of two ACh molecules is required to effect channel opening, however, there is no cooperativity in binding ACh, though the Hill coefficient is $n = 2.0$ (1.97 ± 0.06) for response. Simply, two-site binding is required for opening. These two sites, one on each of the two α chains have different affinities, a requirement that can be rationalized in energetic, as well as structural terms (15, 25).

Conductance and Shape of the Vestibule

The extracellular domain forms a funnel-shaped vestibule, which extends 54 \AA above the membrane plane (figure 2). The outer diameters are from 74 \AA – 81 \AA , and the solvent accessible surface area of the total extracellular region of the structure is $18,500 \text{ \AA}^2$, of which about 20%, or 3780 \AA^2 , forms the lining of the vestibule. A protein wall $24.5 \pm 1.5 \text{ \AA}$ thick surrounds the entry well, insulating the channel from the effects

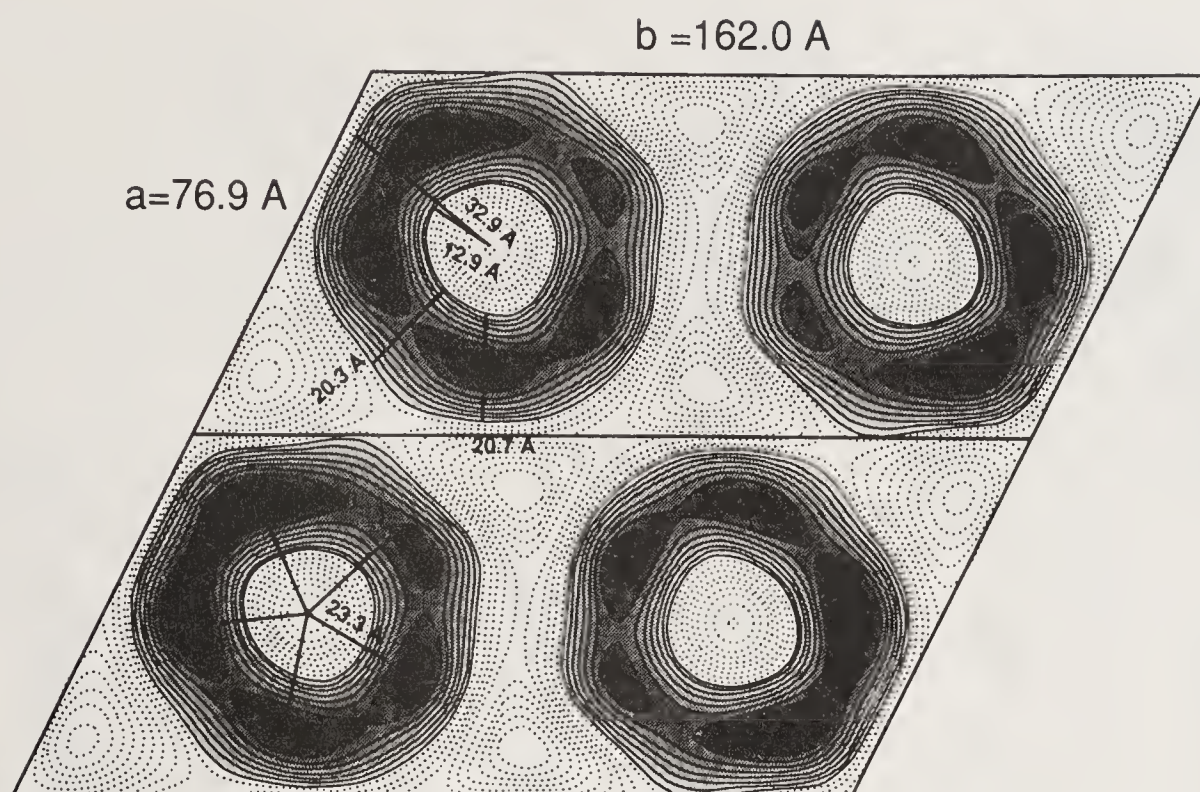


Figure 3. Filtered projection images of alkali-stripped AChR. The pentameric nature of the structure is more apparent after removal of peripheral proteins. The subunit arrangement $\alpha\beta\alpha\gamma\delta$ viewed from the synaptic side of AChR was deduced (15, 23, 24). Solid lines enclose the stain-excluding protein region. Several dimensions of the projected view are shown. The unit cell dimensions for the alkali-stripped form are $a = 76.9 \pm 2.0 \text{ \AA}$, $b = 162.0 \pm 3.2 \text{ \AA}$, included angle $\gamma = 116.0 \pm 1.8^\circ$.

of changes in membrane surface charge. This thickness is typical for the dimensions of an anti-parallel β -barrel structure predicted for the extracellular domain of AChR from amphipathic secondary structure analysis (26) (figure 4).

Cation selectivity and high channel conductance may be due to a negatively charged vestibule which would serve to concentrate ions of opposite charge, and this effect may also be a feature of the super family. The calculated isoelectric point for the extracellular region, aligned sequences¹ 1–228 of the five subunits containing 150 Glu and Asp, 98 Lys and Arg, and 30 His, is 4.77; the net charge expected at pH 7.0 is -50 , and the average charge density is ~ 1 charge per 57 \AA^2 of total solvent accessible surface. Significant in this regard, the GABA and glycine receptors which conduct anions have excess positive charge of about the same amount in their vestibular region, suggesting that charged vestibules may be a general mechanism of enhancing anion or cation selectivity.

Just below the level of the membrane surface leading to the central pore, the diameter of the entry well narrows from 25 to $\sim 7 \text{ \AA}$ on entry to the transbilayer region. Anomalous dispersion X-ray scattering indicates that discrete cation binding sites are found in the *trans*-bilayer region in the resting state (28). The central pore is the gated ion channel, surrounded by a twisted bundle of transmembrane α -helices as schematized in figure 5. On the basis of the observed quasi-five fold symmetry, optimum side chain packing is achieved with a $\sim 12^\circ$ tilt with respect to the axis, and a left-handed twist around the $6.5\text{--}7.0 \text{ \AA}$ pore in the center, which makes the angle between helix axes $\sim 20^\circ$. Constraints on the inner helix packing, at the polar/non-polar interface are probably more stringent than for the external helices, supported by the

¹Sequence numbers in the text refer to a sequence alignment as by Stroud and Finer-Moore (3)

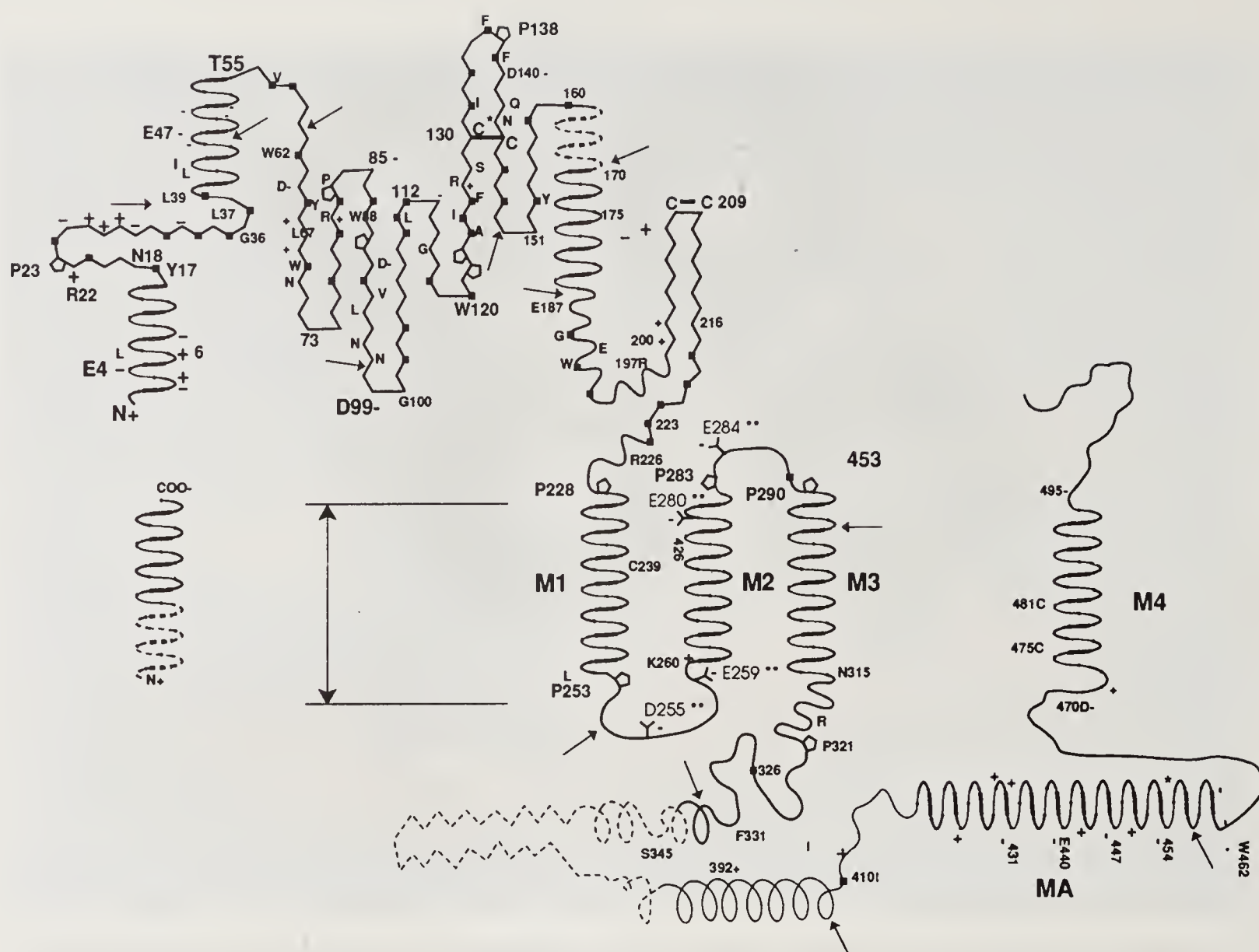


Figure 4. Consensus model of the topology of a neuromuscular AChR subunit, and secondary structure predicted by amphipathic analysis (3,26). The four hydrophobic, putative membrane-spanning α -helices are labeled M1-M4. Residues in M2 whose mutation causes change in conductance are indicated in narrow letters with ** on those that are charged. The conserved cysteines, known to form a disulfide-linked loop in all four neuromuscular AChR subunits C130-C144, enclose the conserved site of N-linked glycosylation (*), found in all neuromuscular AChR subunits and GABA receptor β subunits. The adjacent, disulfide-linked cysteines at the agonist binding site, found only in AChR α subunits are C208 C209. Square blocks indicate conservation of hydrophobic residues in the extracellular side that would tend to fold more toward the inside. Letters identify the more usual residue of highly (not necessarily totally) conserved residues. Positive and negative signs identify regions that generally carry charged side chains. Conserved prolines often occur at the ends of transmembrane helices. Dashed lines indicate variable length substitution between species. Arrows indicate positions of common exon/intron boundaries (26,27).

open structure of some of the hydrophobic helices in contact with the lipid bilayer in the photoreaction-center crystal structure (29).

Noncompetitive Channel Blockers Implicate M2 in Lining the Channel

A binding site for so called 'non-competitive channel blockers' delineates portions of the AChR which form the ion channel. Photolabeling of AChR with the non-competitive blockers trimethisoquin, phencyclidine, histrionicotoxin and chlor-

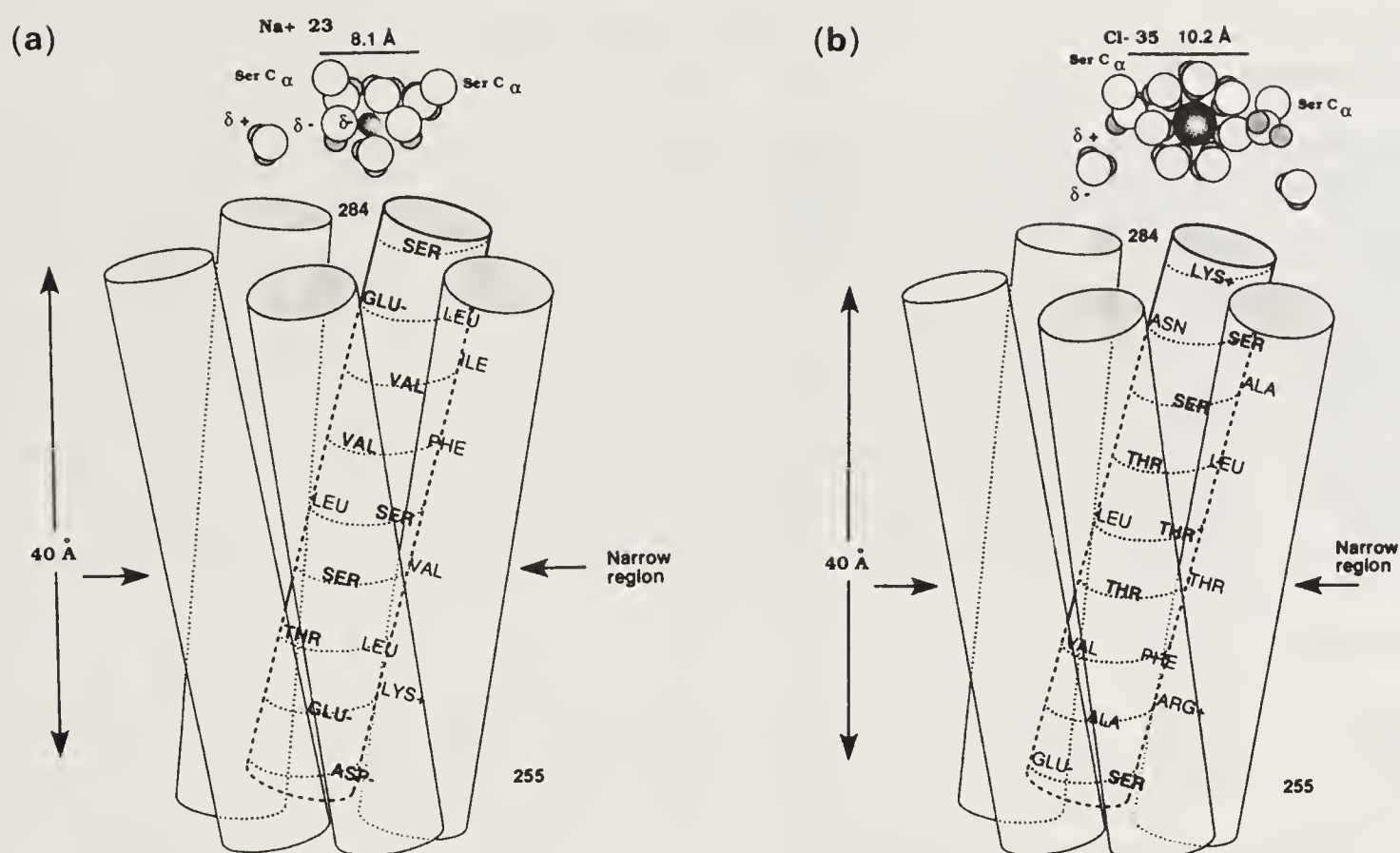


Figure 5. A cut-away depiction of the approximate size and sequence of M2 α -helices packed around a central channel, representing the open state of the channel (a) in cation conducting nicotinic ACh receptor, the α -chain sequence is displayed in precise α -helical configuration, (b) in an anion conducting channel, GABA α -chain is depicted. Reduction of negative charge in (a) at 280 decreased inward current (downward), at 255 it reduced outward current suggesting that these residues may concentrate ions for passage. Changes at 259 had the largest effect on conductance suggesting that the channel is narrowest near 259 (see figure 6). The distances are approximately to scale as shown above, as are the sketches of hydrated sodium (a) and chloride ions (b). Around the ions the inferred different orientation of water molecules and serine side chains are indicated. Hydroxyl groups from serines or threonines within M2 have potential for back-bonding between serine hydroxyl and the carbonyl groups of residues four residues earlier in the sequence, which will present the lone pair of electrons, the δ^- on the oxygen toward the cation channel. The chloride ion is significantly larger, has a smaller energy of hydration, and probably associates with protons on the hydroxyls. The chloride channel sequences are richer in hydroxyl side chains, and threonine rather than serine.

promazine (CPZ) showed incorporation at serine 266 (see figures 5, 6) in M2 (30–33). Thus M2 helices of the individual AChR subunits pack to form the lining of the AChR ion channel, with a non-competitive (open channel) blocker binding site located in the region of the homologous serines 266 (figures 5, 6). Mutagenesis strongly suggests that this model is correct.

M2, then M1, M3, are most Critical to Ion Conductance

The involvement of M2 in ion channel formation is indicated by mutations in M2. Imoto *et al.* (34) showed that charged residues at the ends of M2 affect channel conductance, while Leonard *et al.* (35) showed that the duration of block by the pure open channel blocker QX-222, which only binds to AChR in the open state, decreased monotonically with the number of serine residues removed from the polar side of the

	{255}		{259}		{266}		{270}		{280}		{284}																											
Cation channels																																						
Torpedo alpha	P	T	D	S	G	E	K	M	T	L	S	I	S	V	L	L	S	L	T	V	F	L	L	V	I	V	E	L	I	P	S	T	S	S	A	V	P	
Bovine alpha	P	T	D	S	G	E	K	M	T	L	S	I	S	V	L	L	S	L	T	V	F	L	L	V	I	V	E	L	I	P	S	T	S	S	A	V	P	
mouse alpha	P	T	D	S	G	E	K	M	T	L	S	I	S	V	L	L	S	L	T	V	F	L	L	V	I	V	E	L	I	P	S	T	S	S	A	V	P	
rat neur. a2	P	S	E	C	G	E	K	I	T	L	C	I	S	V	L	L	S	L	T	V	F	L	L	L	I	T	E	I	I	P	S	T	S	L	V	I	P	
Torpedo beta	P	P	D	A	G	E	K	M	S	L	S	I	S	A	L	L	A	V	T	V	F	L	L	L	L	A	D	K	V	P	E	T	S	L	S	V	P	
mouse beta	P	Q	D	A	G	E	K	M	G	L	S	I	F	A	L	L	T	L	T	V	F	L	L	L	L	A	I	K	V	P	E	T	S	L	S	V	P	
Torpedo gamma	P	A	Q	A	G	G	Q	K	C	T	L	S	I	S	V	L	L	A	Q	T	I	F	L	F	L	I	A	Q	K	V	P	E	T	S	L	N	V	P
bovine gamma	P	A	K	A	G	G	Q	K	C	T	V	A	I	N	V	L	L	A	Q	T	V	F	L	F	L	V	A	K	K	V	P	E	T	S	Q	A	V	P
bovine epsilon	P	A	Q	A	G	G	Q	K	C	T	V	S	I	N	V	L	L	A	Q	T	V	F	L	F	L	I	A	Q	K	T	P	E	T	S	L	S	V	P
Torpedo delta	P	A	E	S	G	E	K	M	S	T	A	I	S	V	L	L	A	Q	A	V	F	L	L	L	T	S	Q	R	L	P	E	T	A	L	A	V	P	
bovine delta	P	A	D	C	G	E	K	T	S	M	A	I	S	V	L	L	A	Q	S	V	F	L	L	L	I	S	K	R	L	P	A	T	S	M	A	I	P	
mouse delta	P	G	D	C	G	E	K	T	S	V	A	I	S	V	L	L	A	Q	S	V	F	L	L	L	I	S	K	R	L	P	A	T	S	M	A	I	P	
Dros. ARD	P	A	E	A	G	E	K	V	T	L	G	I	S	I	L	L	S	L	V	V	F	L	L	L	V	S	K	I	L	P	P	T	S	L				
RAT GLU rK	D	P	L	A	Y	E	I	W	M	C	I	V	F	A	Y	I	G	V	S	V	V	L	F	L	V	S	R	F	S	P	Y	E	W	H	S	E	E	
Anion channels																																						
GABA alpha	N	R	E	S	V	P	A	R	T	V	F	G	V	T	T	V	L	T	M	T	T	L	S	I	S	A	R	N	S	L	P	K	V	A	Q	Y	A	T
GABA beta	N	Y	D	A	S	A	A	R	V	A	L	G	I	T	T	V	L	T	M	T	T	I	S	T	H	L	R	E	T	L	P	K	I	P	Y	V	K	
GLYCINE r	M	M	D	A	A	P	A	R	V	G	L	G	I	Y	T	V	L	Y	M	T	T	Q	S	S	G	S	R	A	S	L	P	K	V	S	Y	V	K	

Figure 6. Sequences of M2 regions in cation conducting, ACh, and glutamate activated channels, and in anion conducting, GABA and glycine activated channels illustrate the higher number of polar residues in the anion conducting channels. Mutation of residues at consensus sequence numbers 255, 259, 266, 280, 284 affect conductance (34), as do serines at 266, 270 (35), all of which are expected to lie on the same α -helical face. The polar groups at 259 have the greatest effect on conductance, suggesting that this is the narrowest region within the channel. Rings of charged residues on either side seem to determine conductance from that side.

M2 α -helix. In all four subunit types, there are positively and negatively charged amino acids at both ends of M2 (figures 6, 7). Symmetrical packing of M2 from each subunit to form the lining of the ion channel would place these charged residues in rings at each end (figure 5). For residues 280 on the extracellular entry side, and 255 and 259 on the cytoplasmic side especially, replacement of glutamate, aspartate or glutamine, by glutamine or lysine, in various combinations for all four *T. californica* subunits, led to a decrease in channel conductance in a roughly linear dependence on the decrease in net negative charge (34). Mutations that reduced the total negative charge of the outer rings at either side, 280 or 255, reduced the flow of cations especially from that side, producing rectification, and supporting a model in which rings of negative charge at each end of M2 attract cations to the channel entrance from that side. Changes at 259 produce a larger alteration in the conductance than do changes at either 255 or 280, but do not produce rectification. It is as if 259, near the cytoplasmic surface lies at the narrowest portion of the helix bundle. This is consistent with measurements of streaming potentials which show that the narrowest region is short (36). Replacement of glutamate or aspartate by lysine, or conversely lysine or arginine by glutamate in other portions of the AChR including MA had no effect on channel conductance, though sometimes they impaired expression of AChR. Replacement of serines 266, 270 by alanine at sites within M2 of mouse α , β , and δ subunits also led to irregular rectification and much diminished outward current (35). These results suggest that these serine residues of the sides of M2 that face the center of the channel are important in defining the conductive pathway, and that serine 266 residues near the cytoplasmic end of the putative M2 ion channel may form the non-competitive blocker binding site, at the narrowest region of the open channel (figures 6, 7).

Peptidic Channels

Peptides provide another powerful means for testing the effects of alterations in chemical sequence and structure upon conductance. With the finding that serine-containing M2 regions are implicated in the formation of the ion channel, peptides that match the sequence of M2 were synthesized and presented to membrane bilayers. Such peptides generate ion conducting channels (37). However they lead to a continuum of conductances from 0 to 50 pS. Only a small fraction of openings show a conductance close to that of the AChR (40 pS) in the absence of Ca^{++} ; thus the resultant peptide channels bear little resemblance to the physiologic channels of the parent AChR molecule. Sequences of similar length to M2 but containing only leucines interspersed with serine do form discrete cation conducting channels when there are alternately 2 and 1 serines per turn, but form proton channels when there is only 1 serine per turn of helix, probably due to the smaller number of bundled helices per channel in the less amphipathic sequence (38). At least one, and probably many strongly amphipathic sequences (including one that matches MA) with lysine and carboxylic acids of α -helical period and membrane spanning length (~ 20 amino acids) form very stable, long lived channels of discrete conductances (39). Thus charged residues do not seem to impede ion translocation by binding ions tightly within the pore as might be expected. Peptidic channels are especially useful as a model for mutational effects on conductance, but peptidic models clearly do not resemble the physiologic parent channel.

Gating of the Channel

Evidence from tritium exchange is against any significant change in secondary structure upon agonist induced opening (40), electron microscopy shows that a small degree of twisting seems to accompany desensitization (24). However a major change in the transmembrane region seems to be signalled by the large change in accessibility of all subunits in their transbilayer region to the photoactivable diazirine TID upon activation (40–42). Thus whatever the mechanism by which the signal reaches the channel, a large alteration is seen in the packing of the transbilayer helix bundle. Subunit twisting and allosteric control over structural or electrostatic gating mechanisms within the narrow transbilayer region, felt throughout the bundle, remain viable possibilities.

Characteristics of the AChR Family

All neuronal and neuromuscular junctional AChR subunits have several common features. Most striking are the four putative membrane-spanning stretches of hydrophobic amino acids, M1, M2, M3, and M4 of 27, 20, 20, and 19 amino acids length, respectively (8, 10, 11). As presumed α -helices they would be 40.5 \AA – 29.5 \AA in length, long enough to span the hydrophobic part of the bilayer, which is 41 \AA between head groups (12, 21). Of all the sequence domains of the AChR subunits, these hydrophobic stretches show the highest sequence homology both between subunits, and between species.

The amino terminus is extracellular, and signal sequences are cleaved from the nascent subunits. A highly conserved disulfide bonded loop between Cys 130–Cys 144 (43) with sequence

W-P--I-C---V--FPFD-QNCT/S

is found near the ligand-binding site. There is only one conserved site of N-glycosylation in all species of neuromuscular (but not neuronal) subunits, at N143 in this loop (44, 45). A conserved proline-cysteine sequence is found in M1, and all of the AChR α subunits sequenced thus far have two adjacent cysteines, C208, C209 at or near the agonist binding site that occurs on α chains alone (figure 3).

M1-M3, closely spaced in the sequence, are followed by a long cytoplasmic domain which contains characteristic serine and tyrosine residues that are known sites of phosphorylation (46), and M4 at the carboxy terminus. Common to all sequences of AChR subunits, but not found in anion conducting, GABA or glycine activated channels, is a long sequence of ~ 40 residues, MA, that displays a strong periodic alternation of polar and non-polar residues, with the periodicity of an α -helix (13, 26). This strongly suggests that the sequence lies as an α -helix in an interface between polar and nonpolar environs, such as the membrane or protein surface (figure 3); mutagenesis suggests that this sequence may be important for correct expression of AChR (34).

Permuting Subunits, – Tuning Receptors?

Conservation between species identifies several fingerprints characteristic of each subunit type, and first revealed how the nervous system may alter its AChRs during development. Human α (47) and γ subunits (48), calf α (47), β (49), γ (50), and δ (51), subunits, chick γ and δ subunits (52), and mouse α (53), γ (54), δ subunits (55) have been cloned and sequenced. An additional calf subunit type, called ϵ , was cloned and sequenced and found to be most homologous to the γ subunit (56). The ϵ subunit replaces the γ subunit during calf muscle development and is therefore thought of as the adult form of the γ chain (57, 58). Replacement of calf γ by calf ϵ leads to the expression of AChR with higher conductance characteristic of adult rather than fetal muscle, indicating that properties of AChR are altered during development by changes in subunit makeup. All these sequences show a high degree of homology, both with each other and with their *Torpedo* counterparts, indicating that the neuromuscular AChR has been strongly conserved throughout evolution. In *Xenopus* myotomal muscle there are at least three AChR channel types of different conductances. Two different *Xenopus* α chains which are co-expressed throughout muscle development have been cloned (59). These were found to differ in many regions of the sequence, including M2, in which one has two fewer polar residues than the other again reinforcing subunit substitution as a mechanism for generating diversity in channel conductance.

Many neuronal AChR sequences have been sequenced, and they too suggest a mechanism for diversity through subunit permutation in the central nervous system.

Chloride Conducting Channels, – a More Polar, M2-Derived Pore?

α and β subunits of the bovine γ -aminobutyric acid receptor (GABA receptor) (60) and the strychnine binding (α) subunit of the rat glycine receptor (61) are homologous to the subunits of the AChR throughout their length. Both these chloride conducting channels mediate inhibitory synaptic transmission in the central nervous system. However, the α , β , γ , and δ subunits of the GABA receptor are not thought to form an $\alpha_2\beta\gamma\delta$ complex, analogous to the muscle-type AChR, but form different combinations which are expressed in specific areas of the central nervous system, as are neuronal AChRs. The conserved proline-cysteine in M1 of AChR is found in the GABA receptor α , γ , and δ subunits, and in the form proline-serine in GABA receptor β subunit and glycine receptor α subunit. The common denominator model of an AChR subunit, shown in figure 4, also applies to the subunits of the GABA and glycine receptors, except for the consensus glycosylation site, which is shared only by the GABA receptor β subunits, and the adjacent, disulfide-bonded cysteines near the amino terminal end of M1, which are only found in AChR α subunits. Serines and threonines found in the M2 region of AChR, are found in analogous positions of the M2 sequence of the anion-conducting GABA (60) and glycine receptors (61). Interestingly, the number of serine, threonine, or polar residues is higher, about 2 per turn, than in cation receptor M2s where it is about 1.5 per turn. They are also richer in threonine rather than serine. Serine and threonine side chains frequently back-hydrogen bond to the carbonyls four residues behind in the helix, presenting the δ -charge on the γ -oxygen in toward the channel, especially useful in the cation channels. However similarly oriented threonine side chains would present the γ -CH₃ in toward the channel axis, seeming to reduce its physical diameter. How then could these differences correlate with the larger ionic radius of 1.81 Å for chloride versus sodium ion, 0.95 Å, and the different energies of hydration, which are about twice as high for sodium as for chloride ion? The increased number of polar residues may correlate with a larger number of subunits, as seems to happen in peptidic channels. Alternatively the larger number of water molecules may require the more polar channel seen in the anion conducting sequences. It will be very exciting to see the number and type of subunits required in the anion channels.

Glutamate Receptors in Brain are More Distant Cousins

The main excitatory neurotransmitter in the brain is glutamate. At least three different receptor types of glutamate activated receptors are recognized; these are the N-methyl-D-aspartate (NMDA), kainate, and quisqualate subtypes. Like AChR, they conduct positively charged ions. Hollman *et al.* (62) identified and sequenced a single cDNA sequence that by itself evokes a kainate-type glutamate-activated channel when expressed in oocytes, a selection implied in the cloning strategy used. Though much larger than AChR subunits at 100 kDa, with twice the mass of extracellular protein as the AChR per subunit, it is followed by four hydrophobic stretches, M1, M2, M3, a cytoplasmic domain, and M4. An entire receptor is presumably formed from several copies of the subunit. There is strong conservation in the Cys 130–Cys 143 loop (see

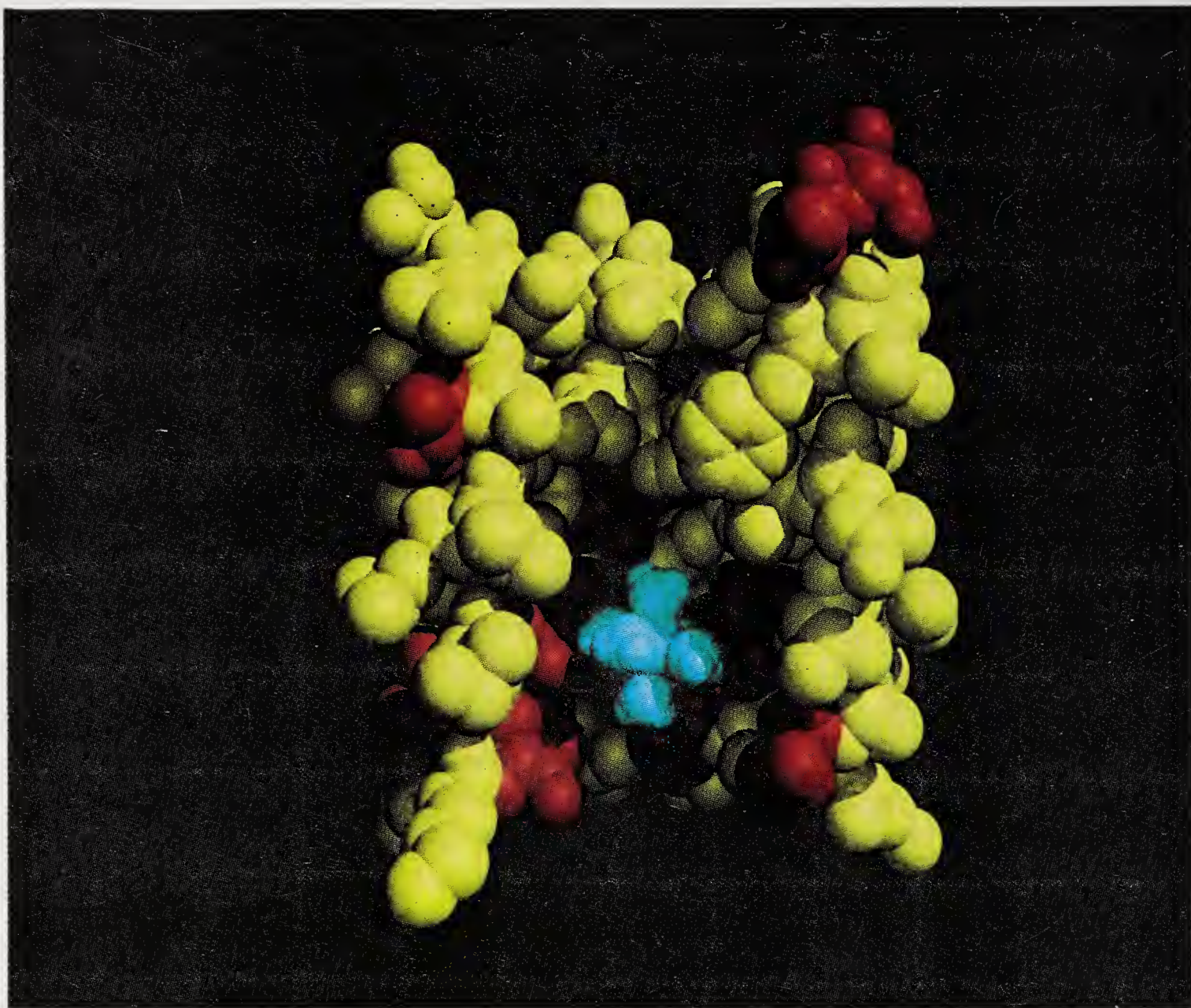


Figure 7(a).

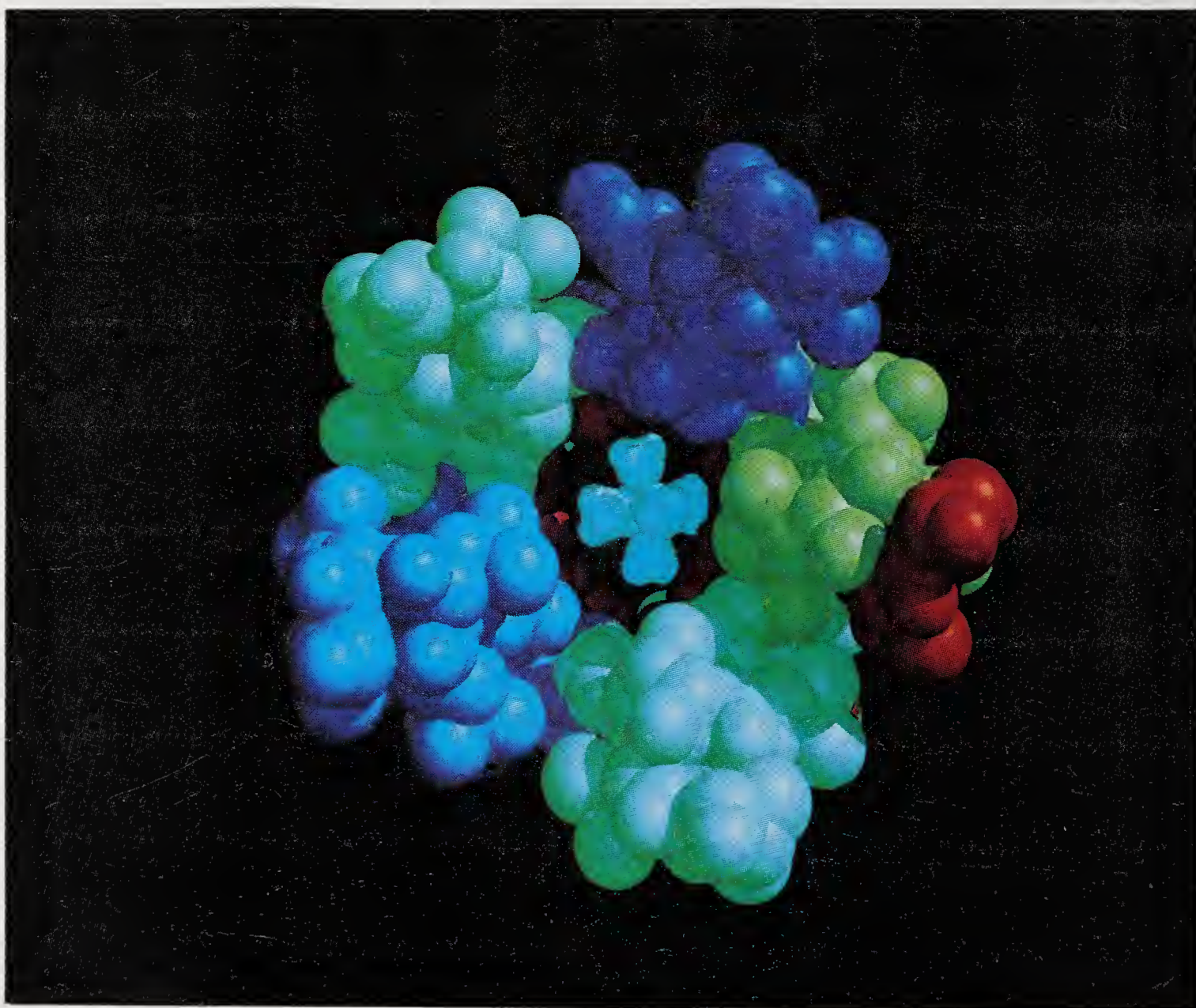


Figure 7(b).

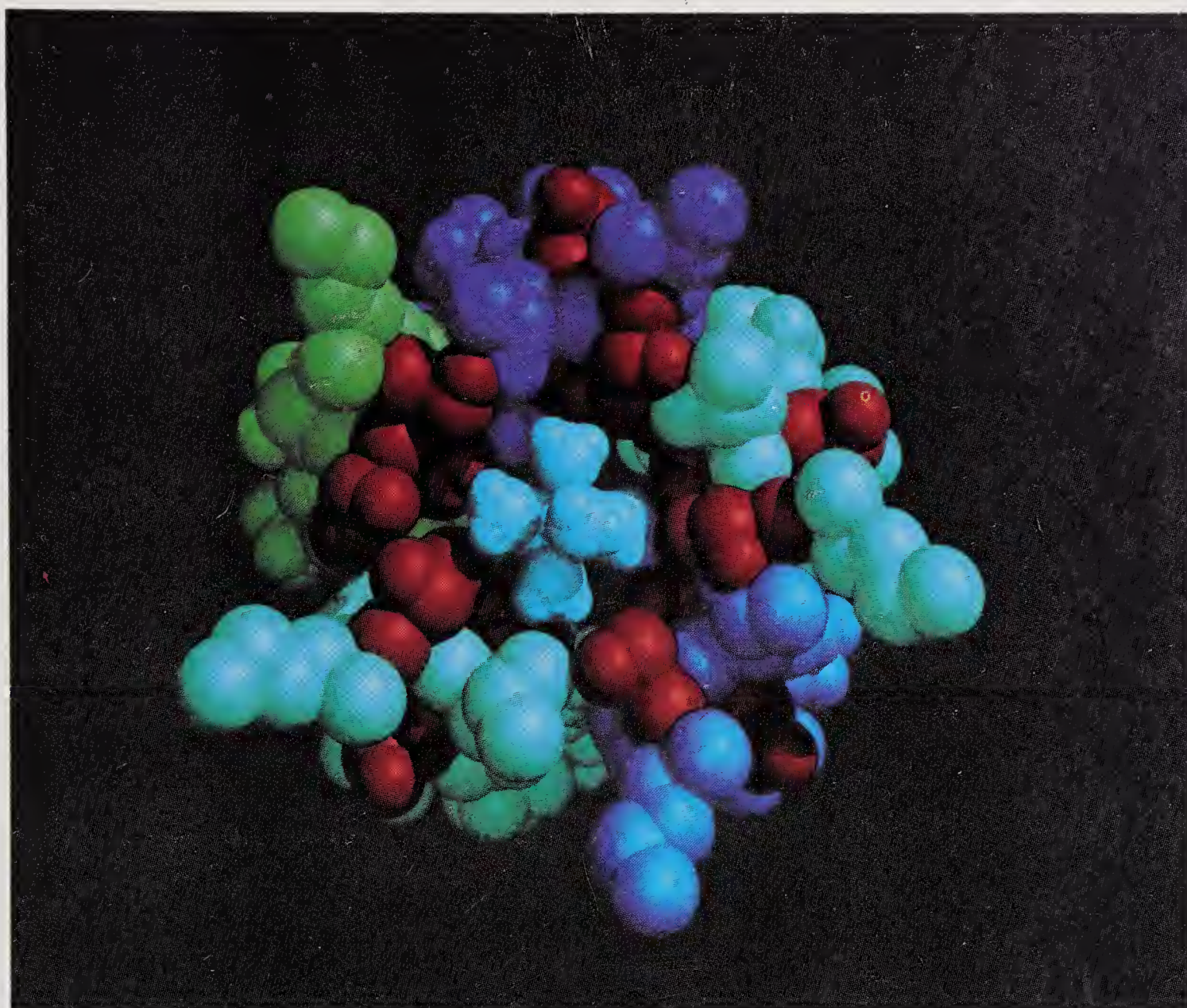


Figure 7(c).

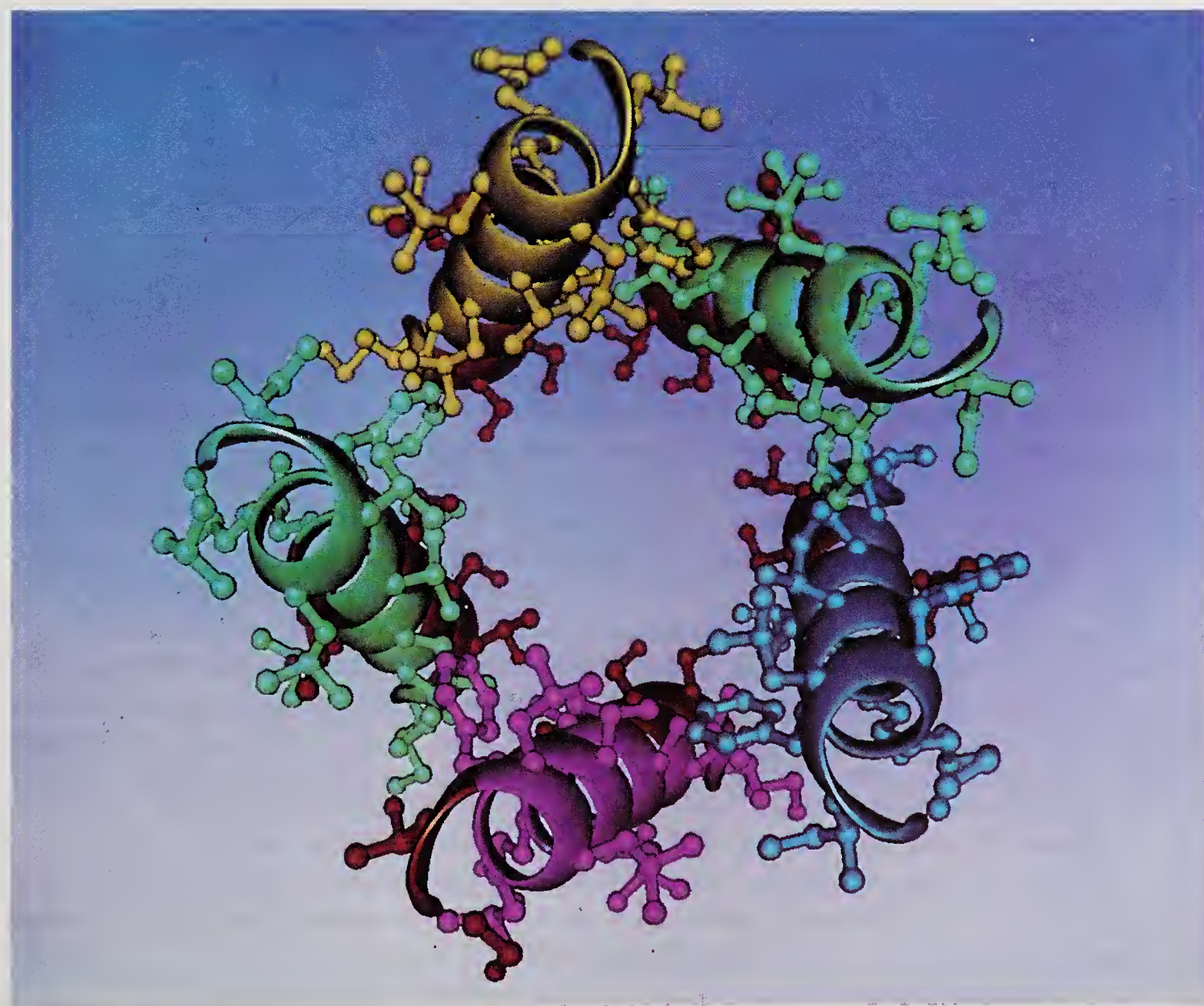


Figure 7(d). (Caption on p. 640).

figure 5 of ref. 62) thus the kainate and nicotinic receptors must be related though the homology elsewhere is slight. The kainate binding protein sequences of Wada *et al.* (63), and Gregor *et al.* (64) are different from each other, and neither is yet proved to form an ionic channel, however both are similar in size to the AChR and both are punctuated by regions analogous to M1, M2, M3 followed by a cytoplasmic domain, and M4.

Topography of Binding Sites and Functional Regions

The neurotoxins have little secondary structure, and are very flexible (65,66). A large part of one side of the 'hand-shaped' toxin molecule is involved in binding AChR (65), probably stretching down into the vestibule from the top crest (figure 8).

Attempts have been made to characterize the antagonist binding site by assessing binding of α -bungarotoxin to small peptides corresponding to regions of the α subunit (*cf* refs. 67–70). Binding peptides span residues 189–220 and include Cys 208, 209 (figure 4).

Toxin binding to AChR evokes a large change in AChR secondary structure of 300 amide N–H groups, compared to very small changes between resting and desensitized states, as assayed by tritium–hydrogen exchange (40). This suggests a mechanism for toxin blockade in which the AChR chains are literally undone, and reform a new β -sheet pattern with the toxin chains. Smaller changes in tritium–hydrogen exchange rates are seen with non-desensitizing, non-competitive local anesthetics (open channel blockers), suggesting that they do not perturb the overall folding of the AChR as they bind within the open channel.

Conformational Change upon Agonist Binding; Desensitization

The structure of ice-embedded *T. marmorata* AChR was redetermined at ~ 20 Å resolution both in the presence and absence of the agonist carbamylcholine. Regions of density extending downward from the top of the AChR, blurring into the less well resolved cytoplasmic domain of the AChR subunits, were slightly differently placed under addition of high (> 100 mM) carbamylcholine, possibly indicating a change in the quaternary structure of the AChR (24). This was assigned to the δ subunit tilting outward from the center, toward the neighboring γ subunit.

Figure 7. Space filling representations of M2 sequences packed together as they may be in the central central, each rotated by 72° about the vertical channel axis, each at the same height, which have quasi five fold symmetry. Packing considerations with these constraints seem to favor a left handed twist to the bundle, and displays a characteristic 'knobs into holes' packing as optimal (3). This is a proposed model which summarizes the current consensus information: **(a)** cross-section through the channel showing serine and other polar residues in red, and a hydrated sodium ion in blue, **(b)** entry well at the synaptic side of the M2 bundle with a hydrated sodium ion, **(c)** bundle of M2 helices seen from the cytoplasmic side. A narrower pore on this side is suggested by effects of mutagenesis on conductance, **(d)** model of the M2 bundle as seen from the synaptic side showing all side chains. Polar side chains within the channel are shown in red.



Figure 8. A ribbon drawing of α bungarotoxin structure determined by Love and Stroud (65), shows it to be an open hand-shaped molecule. Based on chemical labeling and the many related sequences of toxins a large part of the face shown is involved in the interface with AChR, and the large conformational change in AChR seen by tritium-hydrogen exchange may be due to a large refolding which occurs upon toxin binding to AChR, possibly involving reordering of the hydrogen bonding of the putative beta sheet regions that correspond to the binding sequence in AChR. Bungarotoxin may work by interference in the secondary structure of AChR.

The conformation of AChR, as analyzed by tritium-hydrogen exchange, shows no differences between resting and desensitized AChR (40), indicating that the two states differ little in their overall hydrogen-bonded secondary structure and solvent accessibility. A major change in the transmembrane region of all four subunits upon desensitization is seen by agonist-induced alteration in the incorporation of the hydrophobic, photoactivated probe [125 I]TID (41, 42). The largest effect of desensitization upon [125 I]TID labeling is on the γ subunit, whose extracellular domain has also been shown to be perturbed upon desensitization solely on structural grounds

(24). But most importantly, the TID data suggest that ligation of the AChR causes a major change in the configuration of all subunits in their contacts with the lipid bilayer; the helices rearrange during opening.

Perspective

As the best studied member of a super-family of ligand-gated ion channels, AChR provides the most information about the relationship between ion channel structure and function. Much has been learned from the bundled helix model for the ion conducting channels. But much remains unknown about key aspects of ion channel function including the gating mechanism and ion selectivity. Molecular events between ligand binding and channel opening are still undescribed. In addition, a theoretical analysis of the effects of sequence on selectivity and conductivity is now within reach. While a charged vestibule rationalizes gross selectivity, anions versus cations, it does not provide a way to discriminate among mono vs. divalent ions, or among similarly charged ions of different size. These finer types of selection must result from specific interactions between the permeant ions and the channel walls. Crystallization and high-resolution structure analysis by X-ray diffraction of these molecules is still an important goal for furthering our understanding of ion channel function. But the understanding even now has advanced so much in the last two years alone that in the words of C. F. Stevens, "we can anticipate that the next few years will finally provide the data for a rational taxonomy of channels in the brain" (71). And as we do, the conformations of proteins recognized by G. N. Ramachandran play an increasingly large role. The predictive power is very good for the transmembrane portion of membrane proteins, and in terms of such proteins that mediate the transport of material from one side of a membrane to the other, these are the most interesting regions that determine function. Ramachandran's perspective is becoming increasingly invaluable in the interpretation of protein structure and function. Characteristically, the value of his insight is only now, even with the speed of communication in science, becoming recognized as the most powerful descriptor in characterization of protein conformation.

Acknowledgement

I am indebted to G. N. Ramachandran for his insights into protein conformation, first known to me during my graduate studies with J. D. Bernal on small cyclic peptides, in 1968. Images for figure 6 were created by Alexander Kamb and Julie Newdoll, and figures 1 and 7 by Julie Newdoll using software written by Conrad Huang, Eric Pettersen, & Greg Couch at the UCSF Computer Graphics Laboratory. I am grateful to Alexander Kamb for discussion. Figures 2, 3 are reprinted with permission from *J. Cell. Biol.*, figures 4–8 with permission from *Biochemistry*.

References

1. Chambers, J. L. & Stroud, R. M. (1979) *Acta Crystallogr.* **B35**, 1861–1871.
2. McCarthy, M. P., Earnest, J. P., Young, E. F., Choe, S. & Stroud, R. M. (1986) *Annu. Rev. Neurosci.* **9**, 383–413.

3. Stroud, R. M. & Finer-Moore, J. (1985) *Annu. Rev. Cell Biol.* **1**, 317–351.
4. Sumikawa, K., Houghton, M., Smith, J. C., Bell, L., Richards, B. M. & Barnard, E. A. (1982) *Nucleic Acids Res.* **10**, 5809–5822.
5. Ballivet, M., Patrick, J., Lee, J. & Heinemann, S. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4466–4470.
6. Giraudat, J., Devillers-Thiery, A., Auffray, C., Rougeon, F., & Changeux, J.-P. (1982) *EMBO J.* **1**, 713–717.
7. Noda, M., Takashi, H., Tanabe, T., Toyasato, M., Furutani, Y., Hirose, T., Asai, M., Inayama, S., Miyata, T. & Nume, S. (1982) *Nature (London)* **299**, 793–797.
8. Devillers-Thiery, A., Giraudat, J., Bentaboulet, M. & Changeux, J. P. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2067–2071.
9. Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikyotani, S., Hirose, T., Asai, M., Takashima, H., Inayama, S., Miyata, T. & Numa, S. (1983) *Nature (London)* **301**, 251–255.
10. Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikyotani, S., Furutani, Y., Hirose, T., Takashima, H., Inayami, S., Miyata, T. & Numa, S. (1983) *Nature (London)* **302**, 528–532.
11. Claudio, T., Ballivet, M., Patrick, J. & Heinemann, S. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1111–1115.
12. Ross, M. J., Klymkowsky, M. W., Agard, D. A. & Stroud, R. M. (1977) *J. Mol. Biol.* **116**, 635–659.
13. Stroud, R. M. (1983) *Neurosci. Commun. (London)* **1**, 124–138.
14. Kistler, J. & Stroud, R. M. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3678–3682.
15. Kistler, J., Stroud, R. M., Klymkowsky, M. W., Lalancette, R. A. & Fairclough, R. H. (1982) *Biophys. J.* **37**, 371–383.
16. Brisson, A. & Unwin, P. N. T. (1984) *J. Cell Biol.* **99**, 1201–1211.
17. Mitra, A. K., McCarthy, M. P. & Stroud, R. M. (1989) *J. Cell Biol.* **109**, 755–774.
18. Dipaola, M., Czajowski, C. & Karlin, A. (1989) *J. Biol. Chem.* **264**, 15 457–63.
19. Brisson, A. & Unwin, P. N. T. (1985) *Nature (London)* **315**, 474–477.
20. Toyoshima, C. & Unwin, N. (1988) *Nature (London)* **336**, 247–250.
21. Stroud, R. M. & Agard, D. A. (1979) *Biophys. J.* **25**, 495–512.
22. Stroud, R. M. & Finer-Moore, J. (1985) *Ann. Rev. Cell. Biol.* **1**, 317–351.
23. Fairclough, R. H., Finer-Moore, J., Love, R. A., Kristofferson, D., Desmeules, P. J. & Stroud, R. M. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **48**, 9–20.
24. Unwin, N., Toyoshima, C. & Kubalek, E. (1988) *J. Cell Biol.* **107**, 1123–1138.
25. Stroud, R. M., McCarthy, M. P. & Shuster, M. (1990) *Biochemistry* **29**, 11009–11 023.
26. Finer-Moore, J. & Stroud, R. M. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 155–159.
27. Stroud, R. M. (1990) *Biological Mass Spectrometry*, eds. Burlingame, A. & McCloskey, J. (Elsevier Press), pp. 653–670.
28. Fairclough, R. H., Miake-Lye, R. C., Stroud, R. M., Hodgson, K. O. & Doniach, S. (1986) *J. Mol. Biol.* **189**, 673–680.
29. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. (1985) *Nature (London)* **318**, 618–624.
30. Giraudat, J., Dennis, M., Heidmann, T., Chang, J. & Changeux, J.-P. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2719–2723.
31. Oberthur, W., Muhn, P., Baumann, H., Lottspeich, F., Wittmann-Liebold, B. & Hucho, F. (1986) *EMBO J.* **5**, 1815–1819.
32. Giraudat, J., Dennis, M., Hiedmann, T., Haumant, P., Lederer, F. & Changeux, J.-P. (1987) *Biochemistry* **26**, 2410–2418.
33. Giraudat, J., Galzi, J., Revah, F., Changeux, J.-P., Haumont, P. & Lederer, F. (1989) *FEBS Lett.* **253**, 190–198.
34. Imoto, K., Busch, C., Sakmann, B., Mishina, M., Konno, T., Nakai, J., Bujo, H., Mori, Y., Fukuda, K. & Numa, S. (1988) *Nature (London)* **335**, 645–648.
35. Leonard, R. J., Labarca, C. G., Charnet, P., Davidson, N. & Lester, H. A. (1988) *Science* **242**, 1578–1581.
36. Dani, J. A. (1989) *J. Neurosci.* **9**, 884–892.
37. Oiki, S., Danho, W., Madison, V. & Montal, M. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8703–8707.
38. Lear, J. D., Wasserman, Z. R. & Degrado, W. F. (1988) *Science* **240**, 1177–1181.
39. Ghosh, P. G. & Stroud, R. M. (1991) *Biochemistry* **30**, 3551–3557.

40. McCarthy, M. P. & Stroud, R. M. (1989) *Biochemistry* **28**, 40–48.
41. White, B. H. & Cohen, J. B. (1988) *Biochemistry* **27**, 8741–8751.
42. McCarthy, M. P. & Stroud, R. M. (1989) *J. Biol. Chem.* **264**, 10911–10916.
43. Kellaris, K. V. & Ware, D. K. (1989) *Biochemistry* **28**, 3469–3482.
44. Nomoto, H., Takahashi, N., Nagaki, Y., Endo, S., Arata, Y. & Hayashi, K. (1986) *Eur. J. Biochem.* **157**, 233–242.
45. Poulter, L., Earnest, J. P., Stroud, R. M. & Burlingame, A. L. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6645–6649.
46. Huganir, R. L. (1987) *J. Recept. Res.* **7**, 241–256.
47. Noda, M., Furutani, Y., Takahashi, H., Toyosato, M., Tanabe, T., Shimizu, S., Kikuyotani, S., Kayano, T., Hirose, T., Inayama, S. & Numa, S. (1983) *Nature (London)* **305**, 818–823.
48. Shibahara, S., Kubo, T., Perski, H. J., Takahashi, H., Noda, M. & Numa, S. (1985) *Eur. J. Biochem.* **146**, 15–22.
49. Tanabe, T., Noda, M., Furutani, Y., Takai, T., Takahashi, H., Tanaka, K., Hirose, T., Inayama, S. & Numa, S. (1984) *Eur. J. Biochem.* **144**, 11–17.
50. Takai, T., Noda, M., Furutani, Y., Takahashi, H., Notake, M., Shimizu, S., Kayano, T., Tanabe, T., Tanaka, K., Hirose, T., Inayama, S. & Numa, S. (1984) *Eur. J. Biochem.* **143**, 109–115.
51. Kubo, T., Noda, M., Takai, T., Tanabe, T., Kanyano, T., Shimizu, S., Tanaka, K., Takahashi, H., Hirose, T., Inayama, S., Kikuno, R., Miyata, T. & Numa, S. (1985) *Eur. J. Biochem.* **149**, 5–13.
52. Nef, P., Maunon, A., Stalder, R., Allrod, C. & Ballivet, M. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7975–7979.
53. Boulter, J., Luyten, W., Evans, K., Masson, P., Ballivet, M., Goldman, D., Stengelin, S., Martin, G., Heinemann, S. & Patrick, J. (1985) *J. Neurosci.* **5**, 2545–2552.
54. Yu, L., La Polla, R. J. & Davidson, N. (1986) *Nucleic Acids Res.* **14**, 3539–3555.
55. LaPolla, R. J., Mixter-Mayne, K. & Davidson, N. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7970–7974.
56. Takai, T., Noda, M., Mishina, M., Shimizu, S., Furutani, Y., Kayano, T., Ikeda, T., Kubo, T., Takahashi, H., Takahashi, T., Kuno, M. & Numa, S. (1985) *Nature (London)* **315**, 761–764.
57. Mishina, M., Takai, T., Imoto, K., Noda, M., Takahashi, T., Numa, S., Methfessel, C. & Sakmann, B. (1986) *Nature (London)* **321**, 406–411.
58. Gu, Z. & Hall, Z. W. (1988) *Neuron* **1**, 117–125.
59. Hartman, D. S. & Claudio, T. (1990) *Nature (London)* **343**, 372–375.
60. Schofield, P. R., Darlison, M. G., Fujita, N., Burt, D. R., Stephenson, F. A., Rodriguez, H., Rhee, L. M., Ramachandran, J., Reale, V., Glencorse, T. A., Seeberg, P. H. & Barnard, E. A. (1987) *Nature* **328**, 221–227.
61. Grenningloh, G., Rienitz, A., Scjmitt, B., Methfessel, C., Zensen, M., Beyreuther, K., Gundelfinger, E. D. & Betz, H. (1987) *Nature (London)* **328**, 215–220.
62. Hollmann, M., O'Shea, G. A., Rogers, S. W. Heineman, S. (1989) *Nature* **342**, 643–648.
63. Wada, E., Wada, K., Boulter, J., Deneris, E., Heinemann, S., Patrick, J. & Swanson, L. W. (1989) *J. Comp. Neurol.* **284**, 314–335.
64. Gregor, P., Eshhar, N., Ortega, A. & Teichberg, V. I. (1988) *EMBO J.* **7**, 2673–2679.
65. Love, R. A. & Stroud, R. M. (1986) *Protein Eng.* **1**, 37–46.
66. Basus, V. J., Billeter, M., Love, R. A., Stroud, R. M. & Kuntz, I. D. (1988) *Biochemistry* **27**, 2763–2771.
67. Neumann, D., Barchan, D., Safran, A., Gershoni, J. M. & Fuchs, S. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3008–3011.
68. Gotti, C., Mazzola, G., Longhi, R., Fornasari, D. & Clementi, F. (1987) *Neurosci. Lett.* **82**, 113–119.
69. Ralston, S., Sarin, V., Thanh, H. L., Rivier, J., Fox, J. L. & Lindstrom, J. (1987) *Biochemistry* **26**, 3261–3266.
70. Hawrot, E., Colson, K. L., Armitage, I. M. & Song, G. Q. (In Press, 1990) in *Frontiers of NMR in Molecular Biology*, eds. Live, D., Armitage, I. & Patel, D., UCLA Symposia on Molecular and Cellular Biology, New Series, **109**, (Alan R. Liss, Inc., New York, N. Y.).
71. Stevens, C. F. (1987) *Nature (London)* **328**, 198–199.

Molecular Architecture by Using Cyclic Peptides for Ion-Transport Through Lipid Membrane

Yukio Imanishi and Shunsaku Kimura

Department of Polymer Chemistry, Kyoto University, Yoshida Honmachi,
Sakyo-ku, Kyoto 606, Japan

Peptide chemistry has entered an era of creation of new proteins with spatially well-defined structures and the desired functional properties. Although many attempts have been made to synthesize new proteins, it has not been successful until the establishment of site-directed mutagenesis of proteins as well as methodology for elucidation of protein structure. Since protein structure is constituted with several elements of characteristic secondary structures, a rational approach to the synthesis of a new protein is to assemble elemental secondary structures in a desired arrangement. Although elemental secondary structures can be designed with high accuracy, linear polypeptides consisting of these elements will not necessarily fold into the desired arrangement. To overcome this difficulty in constructing a desirable tertiary structure, a novel strategy of using a template molecule has been proposed, in which elemental secondary structures are connected to a template molecule to induce a specific three-dimensional arrangement (template-assembled synthetic proteins) (1).

The template molecule should have a rigid backbone structure in order to restrict the spatial arrangement of peptide elements connected to the template. Cyclic peptides are the most suitable candidates for the template molecule, because cyclic constraint in the molecule reduces the number of available conformational states and limits the orientation of side chains with respect to the cyclic skeleton. These characteristics are very advantageous for the design of three-dimensional structures (2, 3).

Cyclic peptides have been utilized for studies of conformation and mechanism of enzymes, ionophores and peptide hormones (2–7 and references therein). For example, in early investigations of cyclic peptides, conformational analysis was particularly focussed on β - and γ -turns (2, 3). Cyclic peptides have been studied as ionophore models, because cyclic peptides complex with metal ions due to the cooperative coordination of main-chain carbonyl groups to metal ions (2, 3). In addition, cyclic peptides have attracted much attention to peptide hormones, because information on the binding sites of receptors can be obtained by examining the receptor affinities of cyclic analogs of peptide hormones. In the cyclic analogs, the spatial arrangement of side chains and cyclic skeleton are restricted (6).

It should be pointed out that investigations on cyclic peptides have been motivated in the interest of constructing three-dimensional structures with specific functions by using rigid conformations of cyclic skeletons. For example, cyclic peptides cannot

transport metal ions through phospholipid membranes unless a number of carbonyl groups are suitably arranged for cation binding and the cation is efficiently shielded from the external environment by hydrophobic side chains. Similarly, a cyclic analog of a peptide hormone will bind effectively to a receptor only in case that spatial arrangement of side chains fits in to the binding site of the receptor. It can be said that investigation on cyclic peptides is to help develop ways to synthesize proteins. It is therefore worthwhile to understand the basic ideas underlying previous investigations on cyclic peptides. In the present investigation, a series of cyclic peptides are described to show the evolutionary process in the molecular design of peptides for the construction of the intended three-dimensional structure.

Materials and Methods

Materials: Cyclic peptides were prepared by conventional liquid-phase methods. In most cases, a linear peptide with N-hydroxysuccinimide ester at the C-terminal was cyclized in pyridine under high dilution. Usually, the final products were purified by an LH-20 column using methanol as eluant and/or high pressure liquid chromatography (HPLC) using an octadecylated silica-gel column.

Methods: CD spectra were recorded on a J-600 spectropolarimeter (Jasco, Japan). Fluorescence was measured with an MPF-4 or F-3010 spectrophotometer (Hitachi, Japan).

Small unilamellar vesicles were prepared as follows: a dry thin membrane of lipid (usually dipalmitoylphosphatidylcholine (DPPC)) was dispersed in a buffer solution, sonicated and centrifuged at 100,000 g.

Ca^{2+} transport through phospholipid membranes was measured by using DPPC unilamellar vesicles. Dry thin film of DPPC was dispersed in a buffer solution containing Quin 2. After sonication and ultracentrifugation at 100,000 g, DPPC vesicles obtained were eluted through a G-50 column to remove free Quin 2. Quin 2-entrapped vesicles obtained were suspended in a buffer solution containing Ca^{2+} , and fluorescence intensity at 510 nm was measured under moderate stirring (8). The excitation wavelength was 339 nm.

Formation of a hydrophobic pocket in cyclo(L-Lys(Z)-Sar)₄ was examined by fluorescence measurement of 8-anilinonaphthalene-1-sulfonate (ANS). To an ANS (2 μM) solution in buffer (1 mM Tris-HCl, pH 7.6), an ethanol solution of cyclo(L-Lys(Z)-Sar)₄ was added. $\text{Ca}(\text{ClO}_4)_2$ was added in prescribed concentrations, which did not affect the pH of the solution. The excitation wavelength of ANS was 350 nm.

Results and Discussion

Design of Cage for Metal Ions

Valinomycin is a naturally occurring ionophore, which selectively binds K^+ . Analysis of valinomycin/ K^+ complex by X-ray diffraction showed a K^+ sitting inside the cage with a hydrophobic wall and coordinated by six ester carbonyl groups (9). The cage structure of valinomycin indicates several important points for the biological activity as an ionophore:

- i) Selective complexation,
- ii) Large binding constant, and
- iii) Effective shielding of the captured cation from water. The last term facilitates permeation of the complex through a phospholipid membrane. The cage structure of valinomycin/ K^+ complex is stabilized by six intramolecular hydrogen bondings (9).

Cyclo(L-Leu-L-Phe-L-Pro)₄, which has a ring size similar to valinomycin, was synthesized and its complexation with metal ions was investigated (10). The cyclic dodecapeptide binds Ba^{2+} , and conformation of the complex is stabilized by four intramolecular hydrogen bonds (figure 1). However, the complexed cation is not completely included in the pocket and is exposed to an external environment. As a result, the dodecapeptide did not transport Ba^{2+} across the phospholipid membrane. In the early stages of designing the dodecapeptide, it was predicted that the peptide forms a compact C₄-symmetric conformation with β -turns. Thereby, a binding pocket for cation can be formed. However, it was a difficult task to have a cation binding site correctly situated at the center of the hydrophobic cyclic peptide only by changing the primary sequence of a large number of amino acid residues.

However, a three-dimensional arrangement of peptide carbonyl groups can be designed with high accuracy in case that carbonyl groups of side chains as well as main chain are taken into consideration. For example, cyclo(L-Leu-L-Phe-L-Pro)₂ is known to bind Ba^{2+} . Spectroscopic analyses of the complex revealed that the cyclic peptide takes a C₂-symmetric conformation with Ba^{2+} coordinated by four carbonyl groups of the cyclic skeleton (11). It is noticeable that the side chains of Leu residues point to the same side of the cyclic plane as the four carbonyl groups (figure 2a). It was therefore expected that substitution of Leu residues of cyclo(L-Leu-L-Phe-L-Pro)₂ with Cys (Acm) residues results in high cation affinity because of the participation of the side-chain carbonyl groups of Cys(Acm) residues in cation binding (figure 2b). Indeed, the Ba^{2+} binding constant of cyclo(L-Cys(Acm)-L-Phe-L-Pro)₂ was 27-fold of cyclo(L-Leu-L-Phe-L-Pro)₂. NMR measurement showed that the amide carbonyl groups of Cys(Acm) side chains as well as those of cyclic skeleton in cyclo(L-Cys(Acm)-L-Phe-L-Pro)₂ cooperatively coordinated to the cation. These results indicate that a cyclic

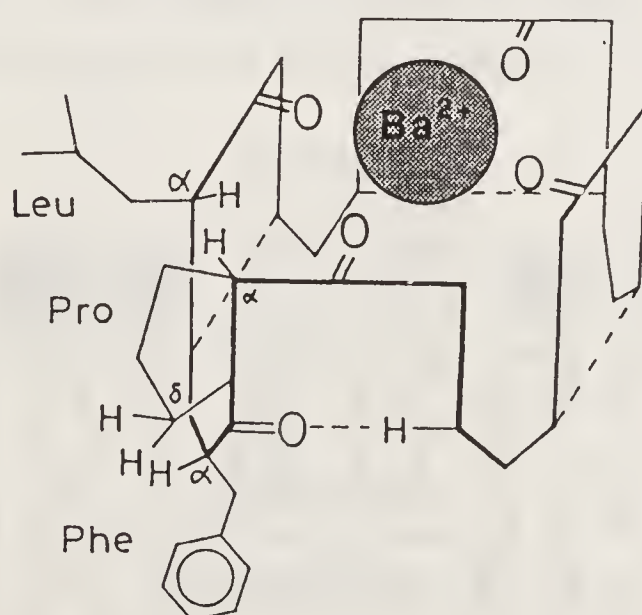


Figure 1. The conformation of cyclo(L-Leu-L-Phe-L-Pro)₄/ Ba^{2+} complex. Ba^{2+} is coordinated by four carbonyl groups, and the structure is stabilized by four intramolecular hydrogen bondings. The ion is exposed to the outside.

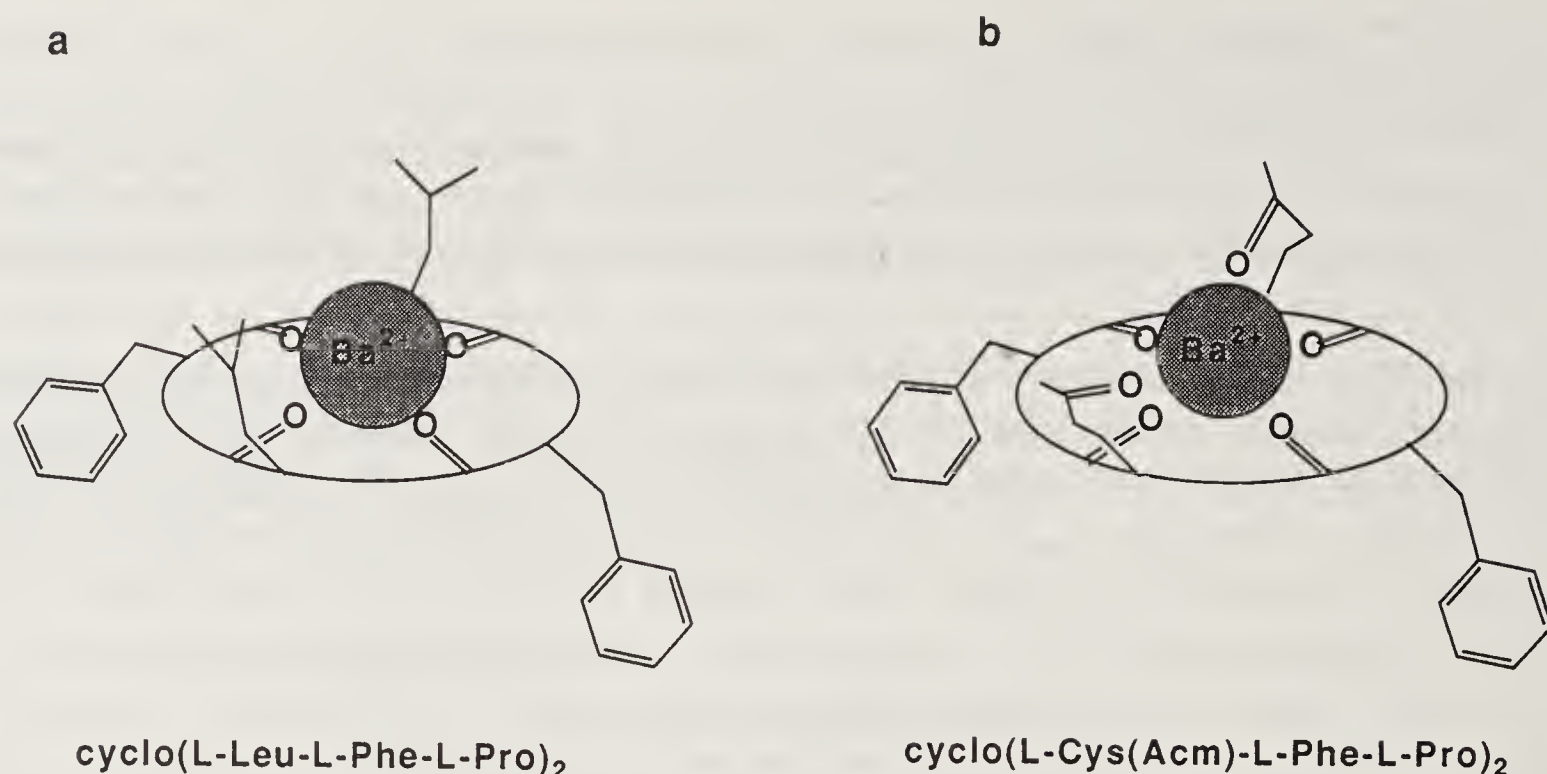


Figure 2. Schematic drawings of Ba^{2+} complex of (a) cyclo(L-Leu-L-Phe-L-Pro)₂ and (b) cyclo(L-Cys(Acm)-L-Phe-L-Pro)₂. Four carbonyl groups of the cyclic skeleton orient to the upper side of the cyclic skeleton and bind Ba^{2+} . The side chain of Leu or Cys(Acm) directs to the same side of Ba^{2+} .

peptide with relatively small ring size can be a suitable template for constructing a desired three-dimensional structure by side-chain orientation.

Amphiphilic Structure and Distribution to Phospholipid Membrane

Some peptide antibiotics and peptide hormones have been shown to interact with phospholipid membranes. It has been pointed out that these membrane-dependent peptides are characterized by their amphiphilic nature, one side of the peptide structure being hydrophobic and the other hydrophilic (12, 13). This type of structure is considered to fit in a membrane surface, because the interfacial region of bilayer membrane with aqueous solution is amphiphilic (figure 3).

Gramicidin S is a cyclic decapeptide antibiotic. It is characteristic of gramicidin S that two L-Orn side chains with positive charges point to one side and hydrophobic L-Val and L-Leu side chains point to the other side of the cyclic skeleton (9, 14). The amphiphilic structure of the peptide should be responsible for strong binding to

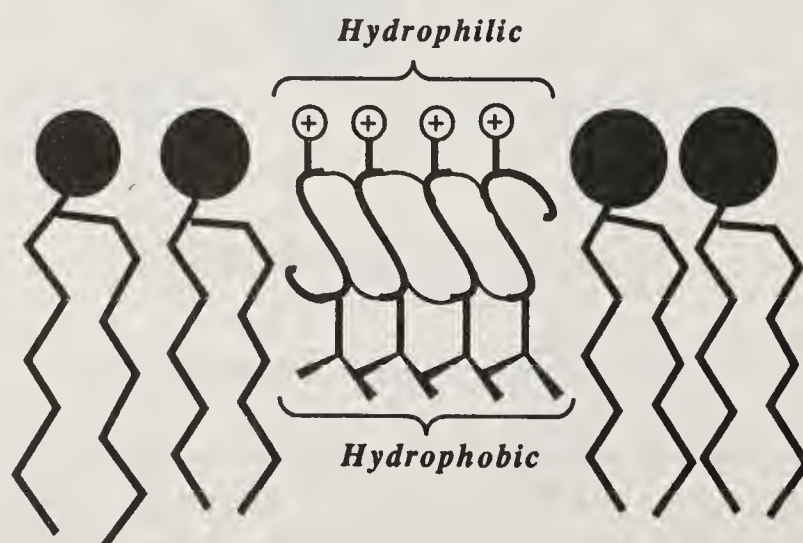


Figure 3. Schematic presentation of the incorporation of an amphiphilic helix into the surface region of a phospholipid bilayer membrane.

phospholipid membranes and the disturbing membrane structure (15). In order to investigate the relation between the amphiphilic structure and the distribution to phospholipid membrane, $\text{cyclo(L-Asp-L-Phe-L-Pro)}_2$ and $\text{cyclo(D-Asp-D-Phe-L-Pro)}_2$ were synthesized (16). These cyclic peptides are constituted by amino acid residues of the same kind but of different chirality. Therefore, any difference in physical properties should be attributed to different conformations. As shown in figure 4, the arrangement of side chains is different in these cyclic hexapeptides. Since two carboxyl groups of $\text{cyclo(L-Asp-L-Phe-L-Pro)}_2$ are distinctly segregated from hydrophobic side chains, it should be more amphiphilic than $\text{cyclo(D-Asp-D-Phe-L-Pro)}_2$.

Change of fluorescence intensity of ANS in the presence of DPPC liposomes was studied with addition of the cyclic hexapeptides (16). ANS has a negative charge and fluoresces intensively when it is bound to the phospholipid membrane. Addition of the cyclic hexapeptides decreased the fluorescence intensity of ANS, indicating that the electric potential of the membrane surface became more negative due to incorporation of negatively charged cyclic hexapeptides to the phospholipid membrane. The effect of $\text{cyclo(L-Asp-L-Phe-L-Pro)}_2$ on decreasing the fluorescence intensity of ANS was greater than that of $\text{cyclo(D-Asp-D-Phe-L-Pro)}_2$. The effect of the cyclic peptides on the structure of phospholipid membranes was examined by IR spectroscopy. While addition of $\text{cyclo(D-Asp-D-Phe-L-Pro)}_2$ did not strongly change the pattern of

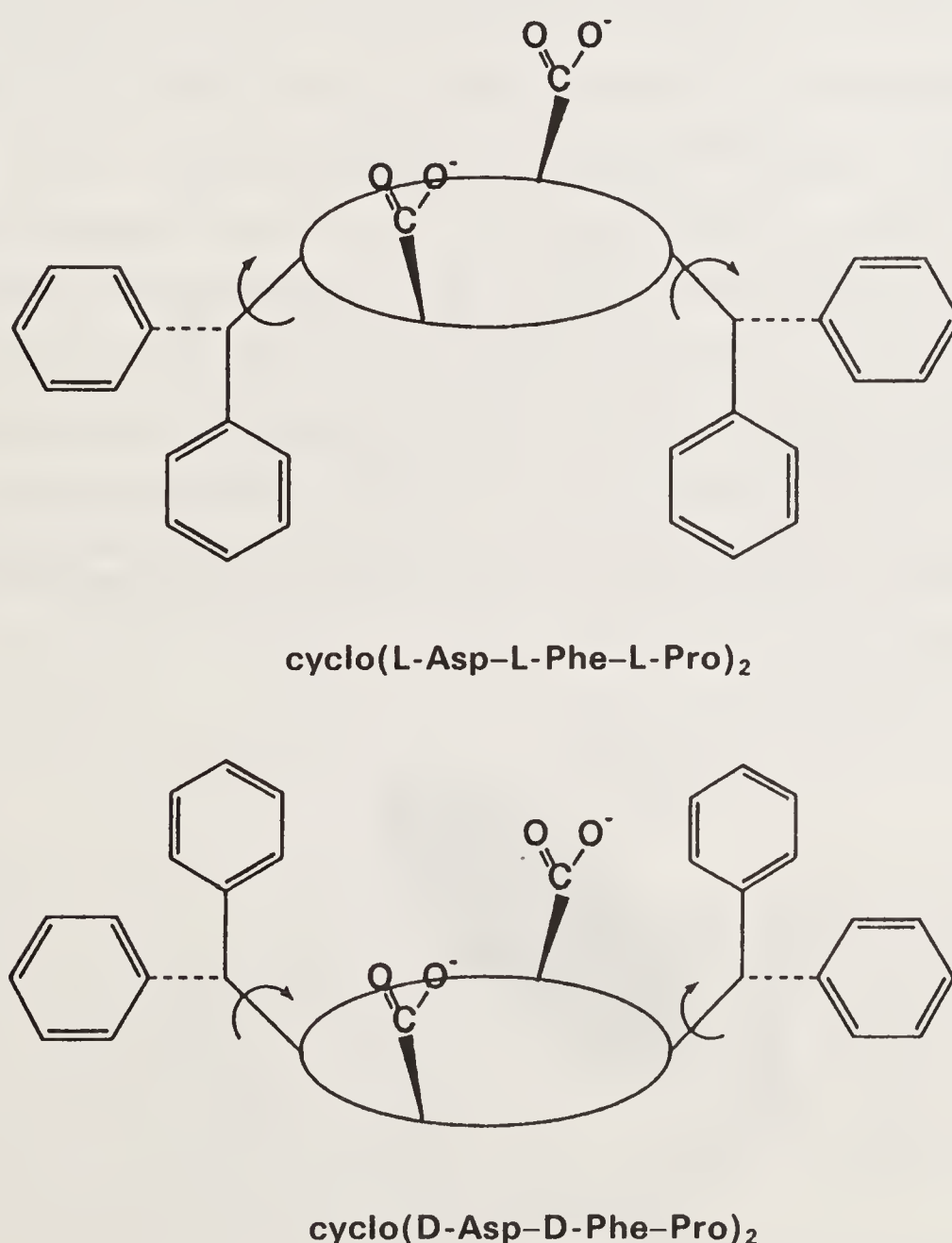


Figure 4. The arrangement of side chains of cyclic hexapeptides. Phenyl and carboxyl groups of $\text{cyclo(L-Asp-L-Phe-L-Pro)}_2$ locate at opposite sides with respect to the plane of the cyclic skeleton, forming an amphiphilic structure.

temperature dependence of absorption due to the symmetric stretching vibration of the phosphate group, addition of cyclo(L-Asp-L-Phe-L-Pro)₂ changed it strongly. The higher solubility of cyclo(L-Asp-L-Phe-L-Pro)₂ into the phospholipid membrane than of cyclo(D-Asp-D-Phe-L-Pro)₂ should be ascribed to the amphiphilic structure of the cyclic hexapeptide.

Cyclic Peptides with Hydrolytic Activity

Enzyme reaction is composed of several steps as shown in figure 5. The first step is the binding of substrate by enzyme, after which the bound substrate is subjected to an intramolecular catalysis. The spatial arrangement of functional groups is elaborated to yield high efficiency in enzymatic reactions. Generally, the activity of a catalytic group is strengthened by other functional groups under intramolecular cooperation (17). In order to mimic enzyme mechanism, cyclic peptides have been used in which functional groups are kept in a desired spatial arrangement to achieve efficient catalysis and substrate specificity. Among them, cyclo(D-Leu-L-Glu-L-His)₂ is unique by showing enantiomer-selective hydrolysis of *p*-nitrophenyl ester hydrochlorides of valine and leucine in the presence of Cu²⁺ (18).

Cyclo(D-Leu-L-Glu-L-His)₂ was about 20 times as efficient as imidazole in catalyzing the hydrolysis of *p*-nitrophenyl laurate in 20% dioxane/water mixture at pH 7.8. The acceleration of hydrolytic reaction by the cyclic hexapeptide is ascribed to hydrophobic binding of the substrate by isobutyl groups of the cyclic peptides. On the other hand, in the case of substrates with lower hydrophobicity such as *p*-nitrophenyl ester hydrochloride of valine and leucine, cyclo(D-Leu-L-Glu-L-His)₂ did not show higher catalytic activity than imidazole. However, upon addition of Cu²⁺ to an aqueous solution of cyclo(D-Leu-L-Glu-L-His)₂, L-enantiomer of the α -amino acid active ester hydrochloride was hydrolyzed faster than the D-enantiomer. It was considered that two carboxyl groups of the cyclic hexapeptide coordinate to Cu²⁺ to induce conformational change. In the complex conformation, the isobutyl group of the Leu residue and imidazolylmethyl group of the His residue are forced to come to the same side of the cyclic skeleton, and the backbone plane of the cyclic hexapeptide and the side chains of Leu and His residues should constitute a hydrophobic pocket (figure

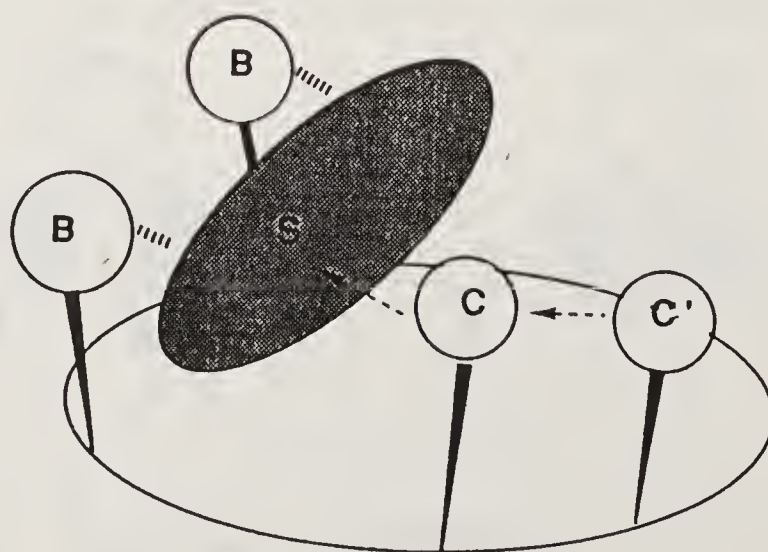


Figure 5. An enzyme model should be constituted with binding site (B) and catalytic site (C). Generally, the activity of the catalytic site is enhanced by another functional group, C'. S represents a substrate.

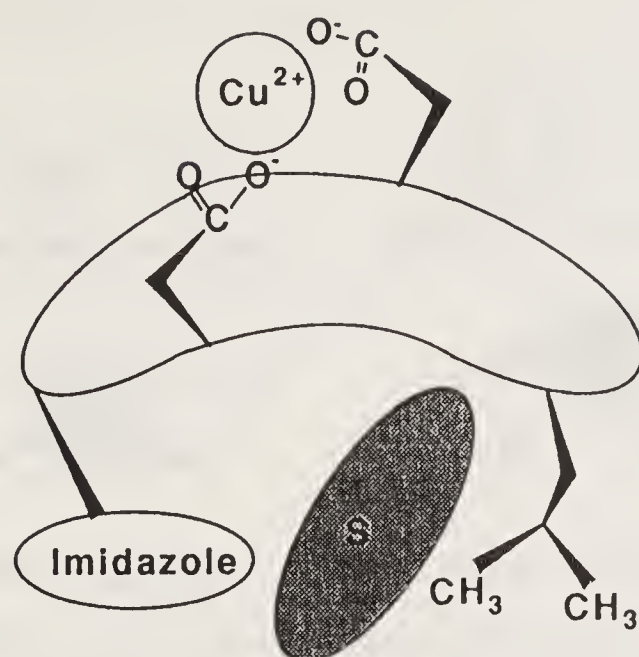


Figure 6. Schematic drawing of Cu^{2+} complex of $\text{cyclo}(\text{D-Leu-L-His-L-Asp})_2$. In the complex, side chains of His and Leu direct to the same side of the cyclic skeleton, forming a binding site for a substrate.

6). The substrate is strongly bound by the hydrophobic pocket and efficiently hydrolyzed. An enantiomer recognition should occur in the substrate binding into the hydrophobic pocket or in the subsequent intramolecular catalysis, resulting in the enantiomer-selective hydrolysis.

Cyclic Peptide with Ionophore Activity

Ionophore translocates cations across phospholipid membranes. Although many cyclic peptides have been reported to form complexes with cations, there have been few examples showing ionophore activity (4). The major reason is the low ability of these cyclic peptides to shield ions from the external environment to enable the complex to permeate through a phospholipid membrane. However, by taking the orientation of hydrophobic side chains into consideration, cyclic peptides can be efficient ionophores in phospholipid membranes.

There are several cyclic hexapeptides which form complexes with Ba^{2+} , in which the carbonyl groups of the cyclic skeleton point with slight tilting toward the outside of the backbone plane and Ba^{2+} is bound by the carbonyl oxygens. These situations are not suitable for permeation through phospholipid membranes. There are at least two ways of designing an inclusion-type complex of cyclic peptides:

- i) Using carbonyl groups in side chains, and
- ii) Increasing the ring size. However, as the ring size increases, the molecular design becomes more difficult. Taking these points into consideration, cyclic octapeptides with highly symmetric molecular structures were extensively investigated. $\text{Cyclo}(\text{Xyz-L-Pro})_4$ was found to form a stable complex with Ba^{2+} and an unstable complex with Ca^{2+} (19, 20) (figure 7a). In order to obtain cyclic octapeptides with a high affinity for Ca^{2+} , which is smaller than Ba^{2+} , cyclic octapeptides with smaller steric hindrances were designed (figure 7b).

Studies on a series of cyclic octapeptides revealed that those consisting of alternating sequences of Sar and a hydrophobic amino acid residue, $\text{cyclo}(\text{Xyz-Sar})_4$, bind Ca^{2+} (21). In order to make the surface of the complex hydrophobic, the orientation of

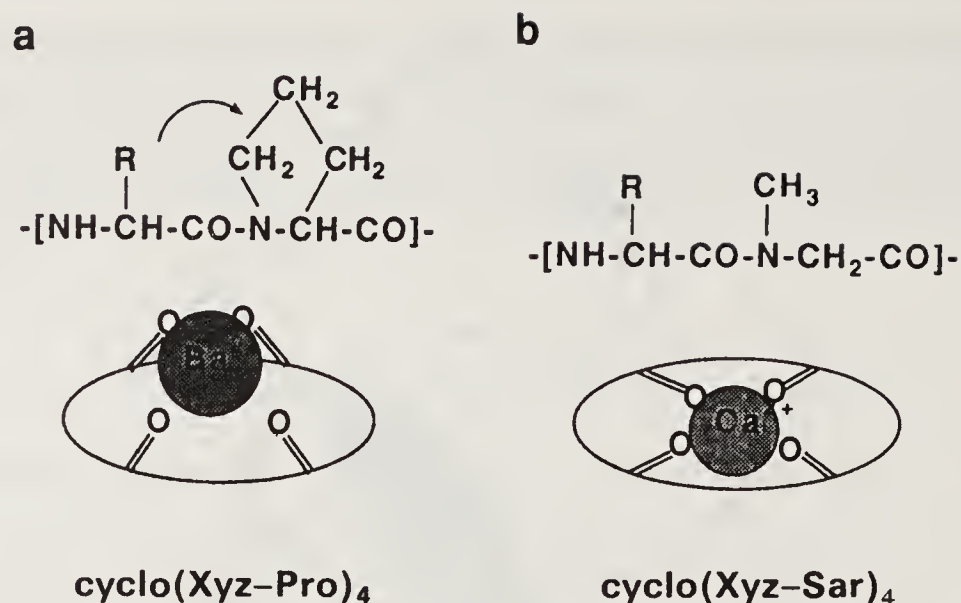


Figure 7. $\text{Cyclo}(\text{Xyz-Pro})_4$ forms a complex selectively with Ba^{2+} . Replacement of Pro residues with Sar residues in the cyclic octapeptide reduces the steric hindrance around the cyclic skeleton, allowing the orientation of carbonyl groups nearly within the plane of the cyclic skeleton and formation of a complex with Ca^{2+} .

hydrophobic side chains was adjusted. Two of the four hydrophobic side chains were directed to one side of the cyclic skeleton and the other two to the opposite side (21). This arrangement can be achieved by combining L- and D-amino acid residues, for example, in $\text{cyclo}(\text{L-Lys}(\text{Z})\text{-Sar-D-Leu-Sar})_2$ (figure 8). As expected, this cyclic octapeptide extracts Ca^{2+} from an aqueous solution to a CHCl_3 solution more efficiently than $\text{cyclo}(\text{L-Lys}(\text{Z})\text{-Sar-L-Leu-Sar})_2$. In the latter, since all hydrophobic side chains point to one side of the cyclic skeleton, the other side is exposed to the aqueous phase (figure 8).

Ca^{2+} transport through a DPPC membrane mediated by the cyclic octapeptides was monitored by measuring the change in fluorescence intensity of Quin 2 encapsulated in the DPPC vesicles. Addition of the cyclic octapeptides increased the fluorescence intensity above the phase-transition temperature of the membrane (figure 9). The increased fluorescence intensity, namely the enhanced Ca^{2+} transport, by $\text{cyclo}(\text{L-Lys}(\text{Z})\text{-Sar-D-Leu-Sar})_2$ was greater than $\text{cyclo}(\text{L-Lys}(\text{Z})\text{-Sar-L-Leu-Sar})_2$ at the same peptide concentration. This result agrees with the water-to- CHCl_3 extraction of Ca^{2+} , indicating that the surface hydrophobicity of the complex governs the transport process across the phospholipid membrane by the cyclic octapeptide.

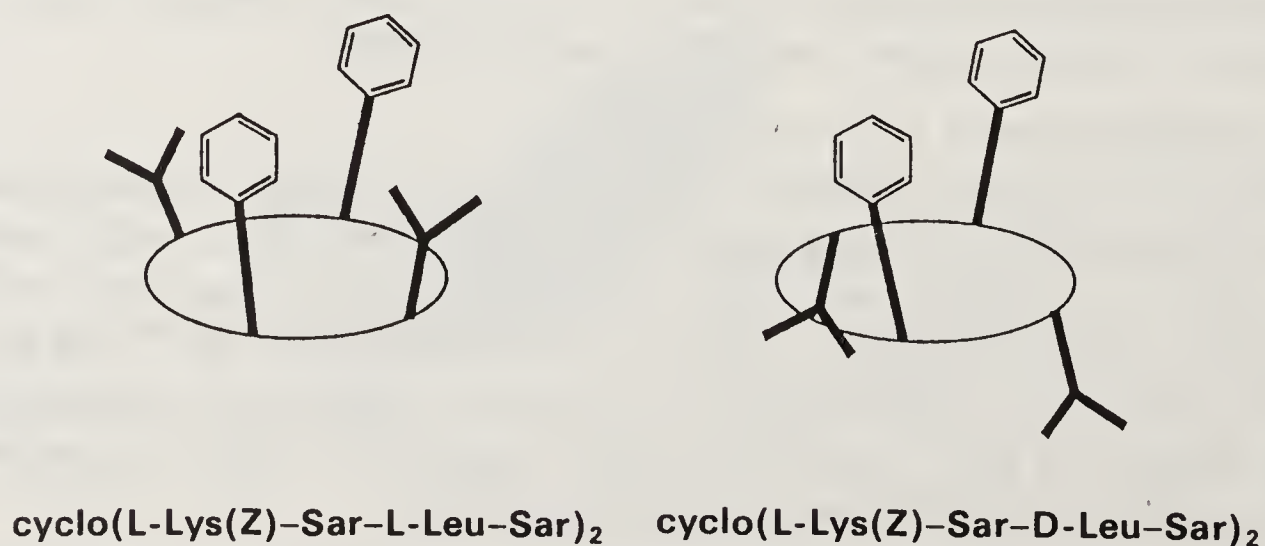


Figure 8. The arrangement of side chains of $\text{cyclo}(\text{L-Lys}(\text{Z})\text{-Sar-L-Leu-Sar})_2$ and $\text{cyclo}(\text{L-Lys}(\text{Z})\text{-Sar-D-Leu-Sar})_2$. In the latter, both sides of the cyclic skeleton are hydrophobic.

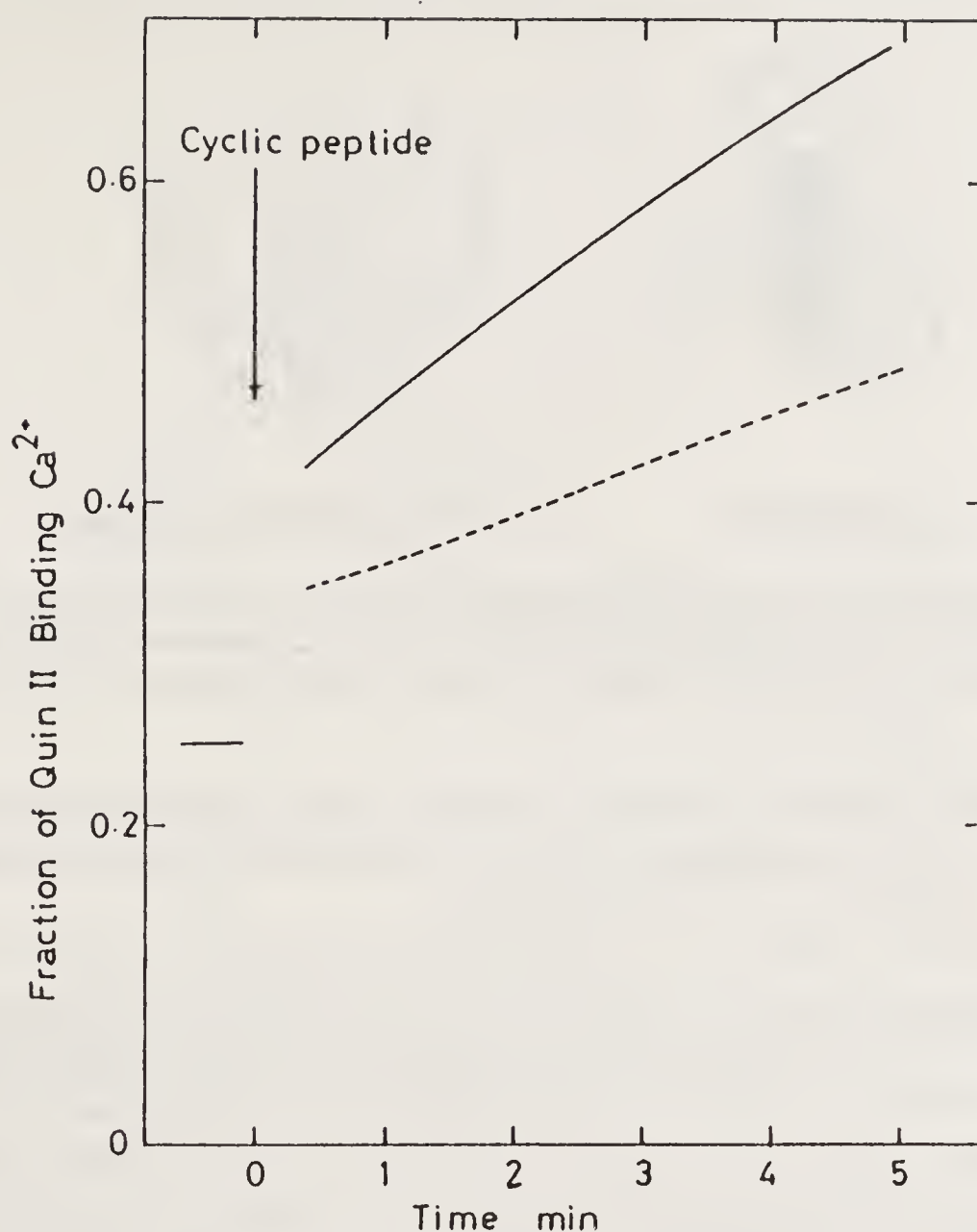


Figure 9. Ca^{2+} transport through DPPC membrane at 49°C by cyclo(L-Lys(Z)-Sar-L-Leu-Sar)₂ ($34\ \mu\text{M}$); — and cyclo(L-Lys(Z)-Sar-D-Leu-Sar)₂ ($5.6\ \mu\text{M}$); ----.

Formation of Hydrophobic Pocket by Cyclic Peptide

Cyclodextrin has been intensively used as an enzyme model for the presence of a hydrophobic pocket which is suitable as a binding site for the substrate (22). The hydrophobic pocket of cyclodextrin is made of relatively rigid constituents, i.e. glucose units. However, such hydrophobic pockets can also be formed by assembling flexible alkyl chains. For example, the octopus-like cyclophane, which is constructed with a rigid macrocyclic skeleton and eight flexible C-14 alkyl chains, provides a cavity for binding hydrophobic compounds in an aqueous solution (23). Cyclodextrin is considered as a model based on the lock-and-key concept, while octopus cyclophane has an induced-fit-type recognition site (figure 10). It has been also pointed out that introduction of additional functional sites into octopus-like cyclophane will be more effective for recognition of the guest molecule (23).

Cyclic peptides with hydrophobic side chains on one side of the cyclic skeleton should form a hydrophobic pocket. This idea is based on either lock-and-key or induced-fit-type concept, depending on whether rigid secondary structures or flexible chains are used for the side chains of the peptides. Advantage of using cyclic peptides, however, is that formation of hydrophobic pockets can be controlled by addition of metal ions, because the conformation of cyclic peptides changes upon complexation.

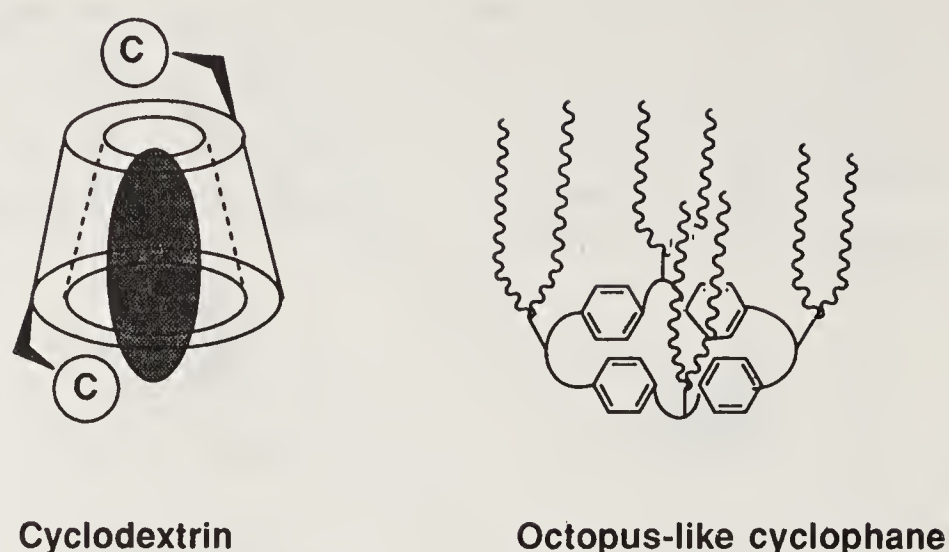


Figure 10. Schematic drawing of the molecular shape of cyclodextrin and octopus-like cyclophane. The interaction of cyclodextrin with a substrate is classified into a key-and-lock type, and that of the cyclophane into an induced-fit type.

For example, cyclic peptides have been shown to mimic metalloenzymes by complexation with metal ions. A hydrophobic pocket is formed or broken upon complexation as described in the section on enzyme models.

Cyclo(L-Lys(Z)-Sar)₄ takes a C₄-symmetric conformation, and four hydrophobic side chains point to the same side of the cyclic skeleton (figure 11a). In order to examine the formation of a hydrophobic pocket by assembly of four Lys side chains of cyclo(L-Lys(Z)-Sar)₄, fluorescence spectra of several hydrophobic probes were measured in the presence of cyclo(L-Lys(Z)-Sar)₄. The fluorescence intensity of ANS in a buffer solution was increased with addition of cyclo(L-Lys(Z)-Sar)₄ (figure 12). Similarly, N-phenyl-naphthylamine and 1,6-diphenylhexa-1,3,5-triene in a buffer solution strengthened the fluorescence intensity by addition of cyclo(L-Lys(Z)-Sar)₄. In the latter case, however, the fluorescent probes seemed to be incorporated into a hydrophobic aggregate of cyclo(L-Lys(Z)-Sar)₄ in a buffer solution. Addition of Ca²⁺ decreased

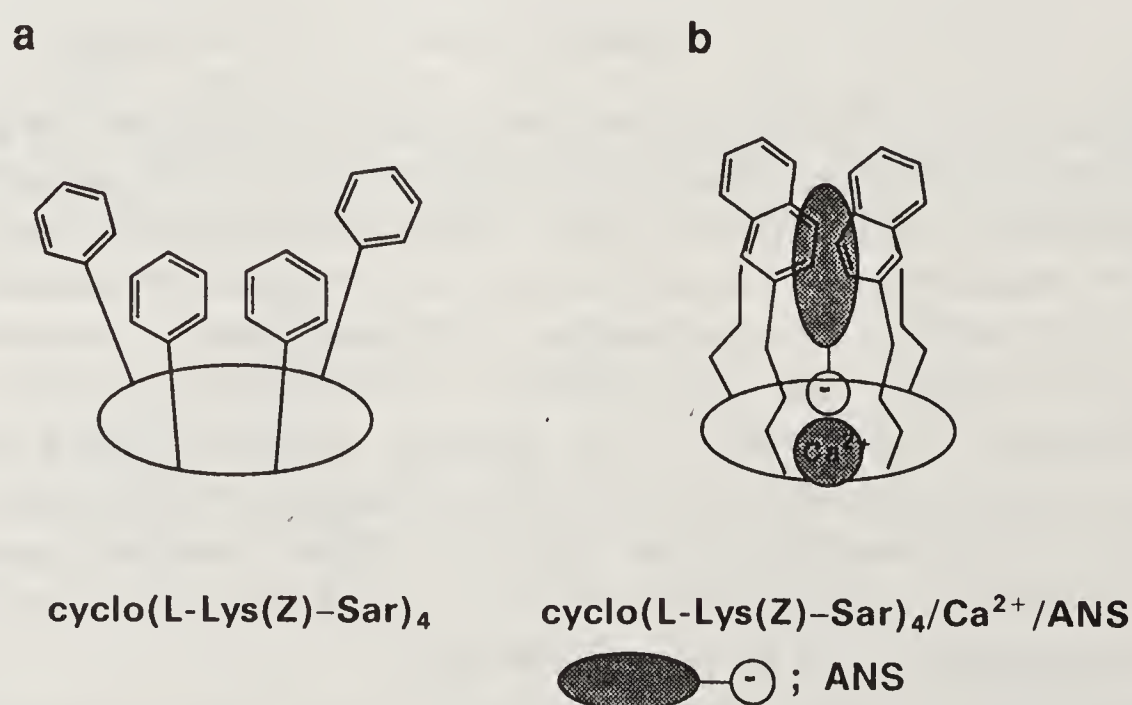


Figure 11. (a) The arrangement of side chains of cyclo(L-Lys(Z)-Sar)₄. (b) Schematic drawing of a ternary complex, cyclo(L-Lys(Z)-Sar)₄, Ca²⁺, and ANS. The binding of Ca²⁺ to the cyclic skeleton induces the formation of a hydrophobic pocket composed of side chains of Lys residues, and ANS is incorporated into the hydrophobic pocket with the aid of interaction of the negative charge of ANS with Ca²⁺.

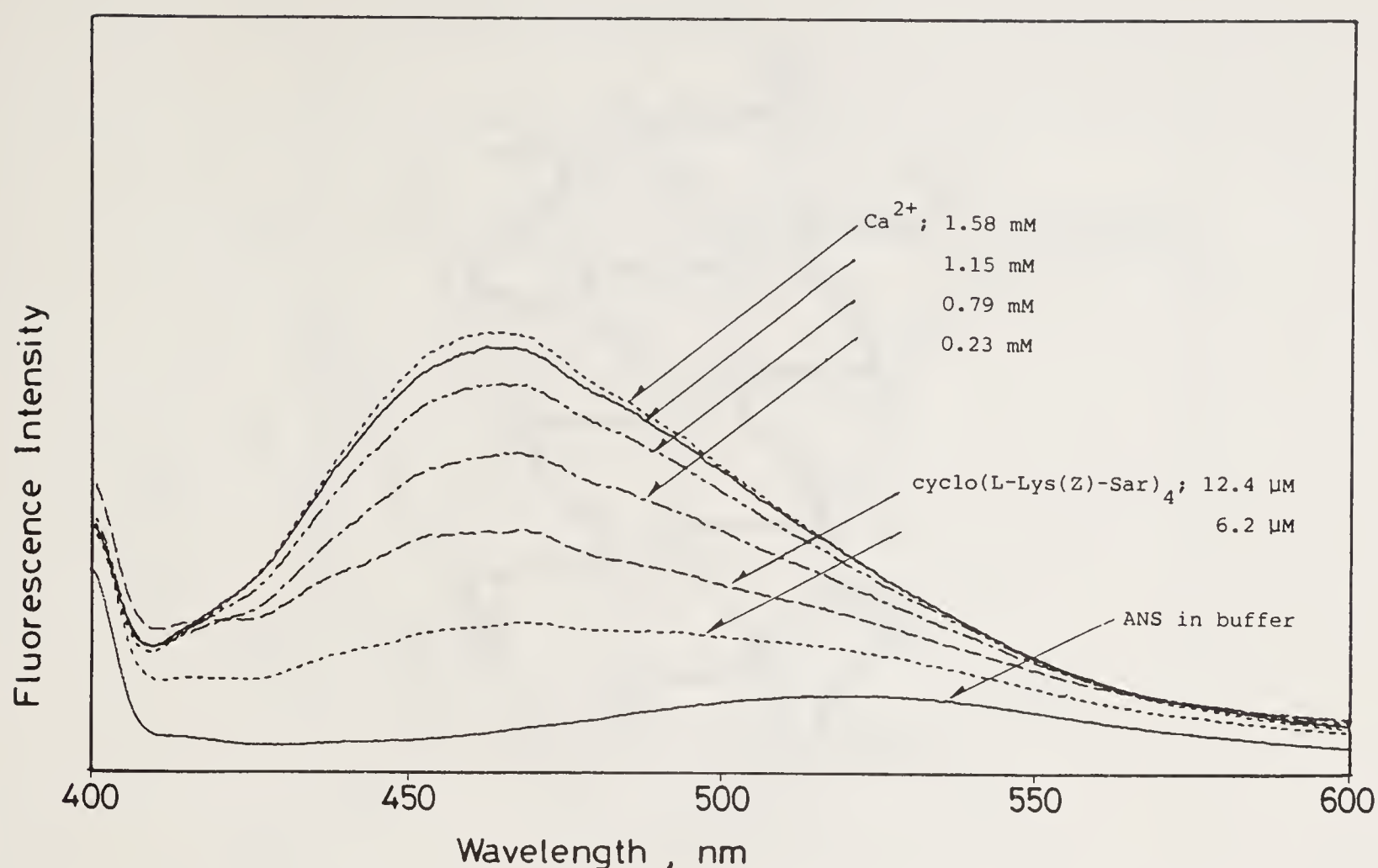


Figure 12. The change of fluorescence intensity of ANS in buffer with the addition of cyclo(L-Lys(Z)-Sar)₄ and successive addition of Ca²⁺.

fluorescence intensity, probably due to collapse of the aggregate as a result of electrostatic repulsion between complexes with positive charges. On the other hand, fluorescence intensity of ANS in the presence of cyclo(L-Lys(Z)-Sar)₄ increased by addition of Ca²⁺ (figure 12). This is probably due to the high stability of a ternary complex of cyclo(L-Lys(Z)-Sar)₄/Ca²⁺/ANS, and ANS should be incorporated in a hydrophobic pocket formed by four Lys side chains (figure 11b).

Regulation of Secondary Structure of Peptide by Ca²⁺

Synthetic peptides with a definite tertiary structure are attractive targets for peptide chemists. An attempt to mimic the three-dimensional structure of antigenic determinant has been reported (24). Generally, antibodies elicited by immunization with excised short sequences of a native protein show low affinity for the native protein, presumably because short linear peptides are unlikely to hold the same conformation as that in native proteins. In order to stabilize α -helical conformation of the 310-327 sequence excised from lactate dehydrogenase C₄, it was transformed to take an $\alpha\alpha$ topology without disturbing the arrangement of solvent-exposing and antibody-contacting residues. An idealized amphiphilic sequence of 18 residues or 40 residues, which takes an $\alpha\alpha$ topology, has been shown to form a stable four-helix bundle (figure 13). The idea of constructing the secondary structure and the supersecondary structure is based on the framework model of protein folding. In brief, α -helices, β -sheets, or turns act as "seeds" for initiation of folding.

On the other hand, Sasaki and Kaiser have succeeded in construction of four-helix bundle by connecting four amphiphilic α -helices to a porphyrin ring, named helichrome (25). Interestingly, the four-helix bundle forms a hydrophobic pocket for

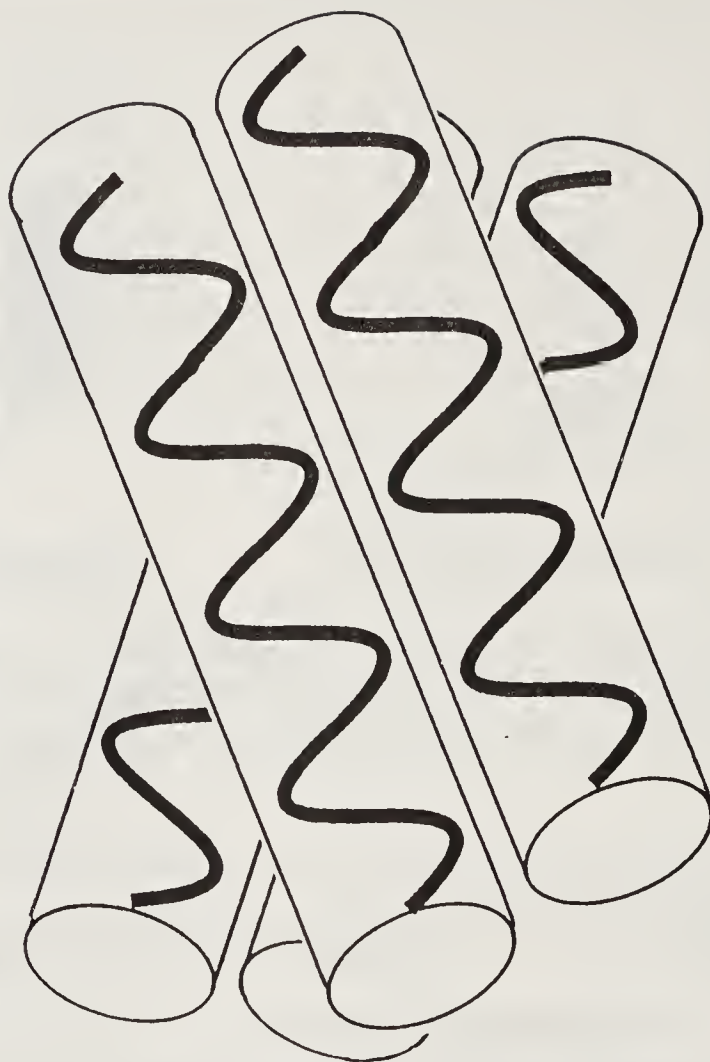


Figure 13. Schematic drawing of the four-helix bundle. The orientation of each helix is antiparallel due to stabilization of macrodipole moments.

substrate binding, and the helichrome showed a hydroxylase activity. Although amphiphilic helices tend to aggregate, the formation of a four-helix bundle of helichrome is brought about by the attachment of the peptides to porphyrin ring.

It will be more significant if the formation of the tertiary structure is regulated by environmental signals. As described in the previous sections, formation of a hydrophobic pocket by assembling the side chains of cyclic peptides is regulated by complexation with metal ions through conformational changes of cyclic skeletons. Accordingly, it is of particular interest to synthesize cyclic peptides with helical peptides as side chains, because secondary structure and supersecondary structure will also be regulated by complexation with metal ions. Noticeably, in this system, the secondary structure of side-chain peptides will also be influenced by interaction between the bound cation and the macrodipole of the helix.

The Boc-(Ala-Aib)₄- group, which is composed of α -helix-forming amino acids and β -sheet-breaking Aib, is connected to each of four side chains in cyclo(Lys-Sar)₄ (C8KS-AA'8). CD measurements of C8KS-AA'8 in 95% methanol reveal that the α -helix content of Boc-(Ala-Aib)₄-chain connected to cyclic peptide is lower than that of Boc-(Ala-Aib)₄-OMe. Presumably, steric hindrance between side chains causes the difference. However, α -helix content of C8KS-AA'8 increases on adding Ca²⁺ in 95% methanol (figure 14). It has been reported that helix content is usually increased by the presence of either the negatively charged N-terminal amino acid residue or the positively charged C-terminal amino acid residue because the charged amino acid residue stabilizes the helical structure due to interaction with the macrodipole of the peptide (26). For example, poly(Lys-Aib-Leu-Aib) was shown to take a helical conformation in an aqueous solution. The helix content was dependent on the pH of

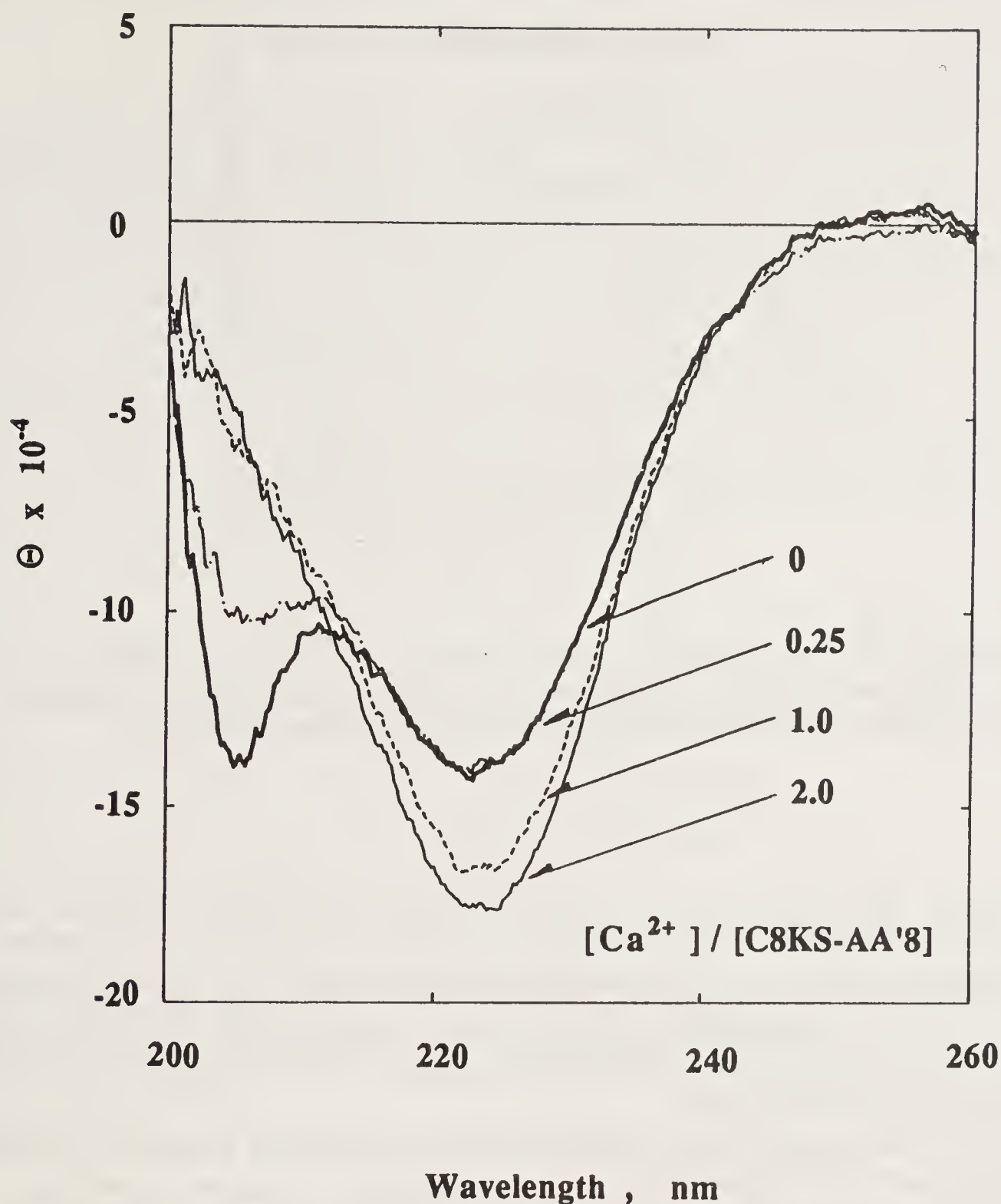


Figure 14. The change of CD spectra of C8KS-AA'8 in 95% methanol with the addition of Ca^{2+} . The intensity of the Cotton effect at 222 nm decreased with the addition of Ca^{2+} , indicating the increase of helix content. The Cotton effect at 208 nm increased, because the Cotton effect arising from the cyclic skeleton exceeded that of the side chains.

the medium, and increased beyond pH 7.1 (figure 15) (27). Since a conformational change should occur upon deprotonation of the N-terminal α -amino group of the peptide, the increased content of helix can be ascribed to elimination of unfavorable interaction of the N-terminal positive charge with the macrodipole of the helix. It is therefore concluded that the increased content of helix in C8KS-AA'8 by addition of Ca^{2+} should be ascribed to favorable interaction of Ca^{2+} with the C-terminal negative macrodipole of the helix (figure 16).

It was expected that C8KS-AA'8 forms hydrophobic pockets composed of four helical segments. However, fluorescence intensity of ANS, phenylanthramine, or 1,6-diphenylhexa-1,3,5-triene did not increase by addition of C8KS-AA'8 and Ca^{2+} . This is in contrast to cyclo(L-Lys(Z)-Sar)₄, which formed a hydrophobic pocket in a buffer solution. In the case of C8KS-AA'8, the macrodipoles of four helical segments

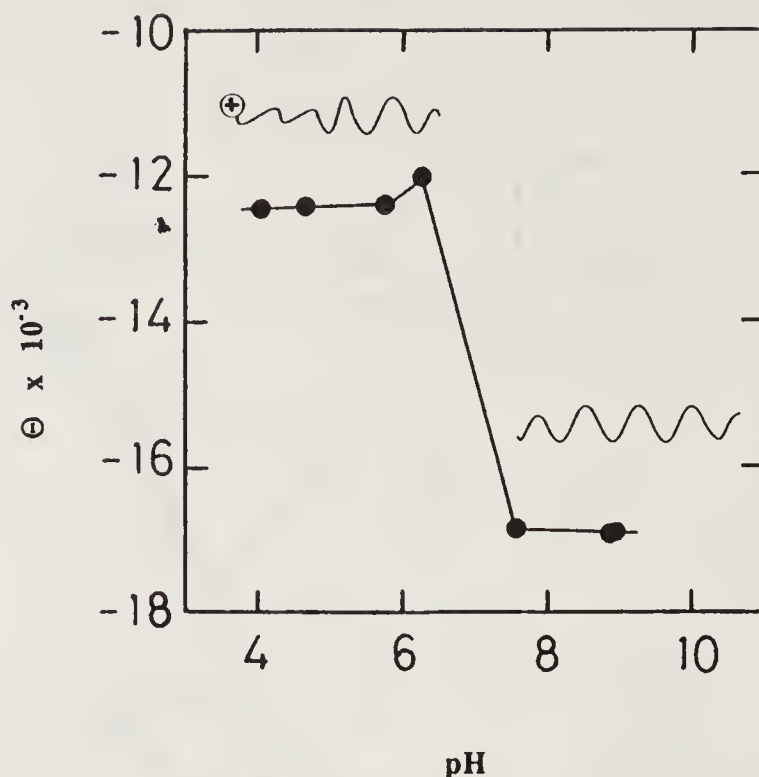


Figure 15. The pH dependence of molar ellipticity at 222 nm of poly(Lys-Aib-Leu-Aib) in water. The N-terminal positive charge reduces the helix content due to the unfavorable interaction with the macrodipole moment of the helix.

might repulse one another. In order to construct a four-helix bundle in a parallel orientation, hydrophobic interactions between helices should be enhanced by elongation of peptide chain or using amphiphilic helical peptides. Indeed, cyclic octapeptide connected with four amphiphilic helical peptides, $\text{Boc}(\text{Lys-Aib-Leu-Aib})_2^-$, provides a hydrophobic pocket upon complexation with Ca^{2+} , showing properties as a model peptide for allosteric proteins.

However, it is expected that C8KS-AA'8 forms four-helix bundle in phospholipid membrane and ion channel across the membrane, because alamethicin forms voltage-dependent ion channels by forming helix bundles. In the case of C8KS-AA'8, part of the cyclic peptide works not only as a template for the helix bundle but also as a selective binding site for the metal ion.

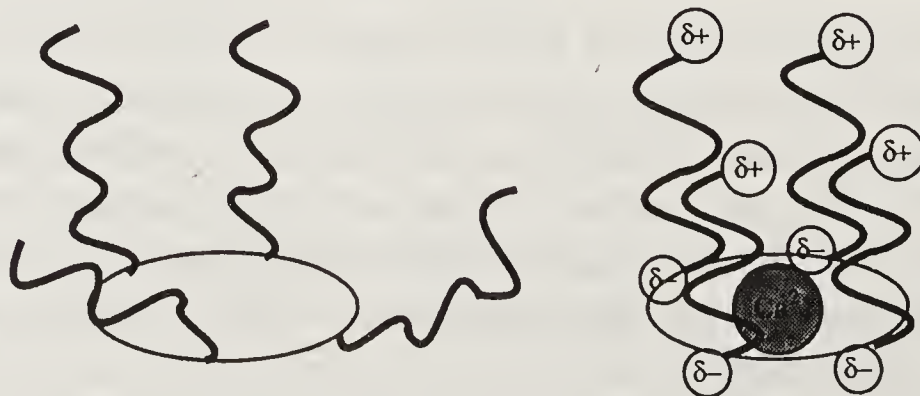


Figure 16. Schematic drawing of the molecular structure of C8KS-AA'8. When the cyclic skeleton binds Ca^{2+} , the helix content of side chains increases due to stabilization of the macrodipole moment.

References

1. Mutter, M. & Vuilleumier, S. (1989) *Angew. Chem., Int. Ed. Engl.* **28**, 535–554.
2. Imanishi, Y. (1976) *Adv. Polym. Sci.*, **20**, 1–76.
3. Deber, C. M., Madison, V. & Blout, E. R. (1976) *Acc. Chem. Res.* **9**, 106–113.
4. Deber, C. M. (1980) *Can. J. Biochem.* **58**, 865–870.
5. Kopple, K. D., Sarkar, S. K. & Giacommetti, G. (1981) *Biopolymers* **20**, 1291–1303.
6. Schiller, P. W. & DiMaio, J. (1982) *Nature (London)* **297**, 74–76.
7. Madison, V. (1985) *Biopolymers* **24**, 97–103.
8. Kimura, S., Ozeki, E. & Imanishi, Y. (1989) *Biopolymers* **28**, 1235–1246.
9. Ovchinnikov, Yu. A. & Ivanov, V. T. (1975) *Tetrahedron* **31**, 2177–2209.
10. Ozeki, E., Kimura, S. & Imanishi, Y. (1988) *J. Chem. Soc., Perkin Trans. II* 1743–1748.
11. Ozeki, E., Kimura, S. & Imanishi, Y. (1989) *Int. J. Pept. Protein Res.* **34**, 111–117.
12. Pownall, H. J., Knapp, R. D., Gotto, A. M. & Massey, J. B. (1983) *FEBS Lett.* **159**, 17–23.
13. Kaiser, E. T. & Kezdy, F. J. (1984) *Science* **223**, 249–255.
14. Schwyzer, R. & Sieber, P. (1957) *Helv. Chim. Acta* **40**, 624–639.
15. Kato, T., Waki, M., Matsuura, S. & Izumiya, N. (1970) *J. Biochem.* **68**, 751–753.
16. Ozeki, E., Miyazu, T., Kimura, S. & Imanishi, Y. (1989) *Int. J. Pept. Protein Res.* **34**, 97–103.
17. Kopple, K. D., Jarabak, R. R. & Bhatia, P. L. (1963) *Biochemistry* **2**, 958–964.
18. Tanihara, M. & Imanishi, Y. (1983) *Polym. J.* **15**, 499–507.
19. Kimura, S. & Imanishi, Y. (1983) *Biopolymers* **22**, 2191–2206.
20. Kimura, S. & Imanishi, Y. (1983) *Biopolymers* **22**, 2383–2395.
21. Kimura, S., Ozeki, E. & Imanishi, Y. (1989) *Biopolymers* **28**, 1247–1257.
22. Komiyama, M. (1988) *Chem. Lett.*, 1121–1124.
23. Murakami, Y., Kikuchi, J., Suzuki, M. & Matsuura, T. (1988) *J. Chem. Soc., Perkin Trans. I* 1289–1299.
24. Kaumaya, P. T., Berndt, K. D., Heidorn, D. B., Trewhella, J., Kezdy, F. J. & Goldberg, E. (1990) *Biochemistry* **29**, 13–23.
25. Sasaki, T. & Kaiser, E. T. (1989) *J. Am. Chem. Soc.* **111**, 380–381.
26. Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M. & Baldwin, R. L. (1987) *Nature (London)* **326**, 563–567.
27. Kono, K., Kimura, S. & Imanishi, Y. (1990) *Biochemistry* **29**, 3631–3637.

Signal Sequences: The Relevance of Their Conformations and Membrane Perturbing Ability in the Sorting of Secretory Proteins

R. Nagaraj

Centre for Cellular and Molecular Biology, Uppal Road,
Hyderabad 500 007, India

Introduction

Extensive research in recent years has revealed that contiguous peptide sequences of ~ 10–40 residues are responsible for targeting of proteins to the inner membrane in bacteria (1, 2), endoplasmic reticulum (ER) in eukaryotes (3, 4) and organelles such as peroxisomes (5), mitochondria (6) and nuclei (7). These sequences, referred to as signal sequences, also have a role in initiating translocation of proteins *across* the *E. coli* inner membrane, ER membrane and mitochondria. Proteins destined for secretion and uptake into mitochondria are initially synthesized as precursors with the signal sequences occurring at the N-terminus (6). Once the targeting and initiation of translocation is achieved, the signal sequences are cleaved off by signal peptidases present in the ER (8), inner membrane of *E. coli* (9) and mitochondrial matrix (10). In the case of proteins destined for the nucleus and peroxisomes, the signals are not removed from the proteins and they do not occur specifically at the N-terminus (5, 7).

In addition to signal sequences, several proteins having a role in secretion have been identified (11–14). Some of them are presumed to interact with signal sequences and play a role in targeting of proteins to organelles or translocation competent membranes, whereas others interact with precursor proteins and maintain them in a 'translocation-competent conformation'. There is remarkable specificity in the targeting of proteins as well as interaction of preproteins with the protein components of the cells' export machinery and this occurs in spite of the absence of primary structure homology amongst signal sequences (15). Although the process of membrane vesicle fusion plays an important role in the processes of intra- and intercellular protein traffic, all non-cytoplasmic proteins have to cross at least one membrane barrier before they reach their final destination. This article focusses attention on the conformational and physico-chemical studies on synthetic signal sequences of secretory protein and, based on these studies, attempts to answer the questions:

- (i) What is the common feature in signal sequences that is recognised by components of the cells' export machinery?
- (ii) How do signal sequences initiate translocation of proteins across membranes?

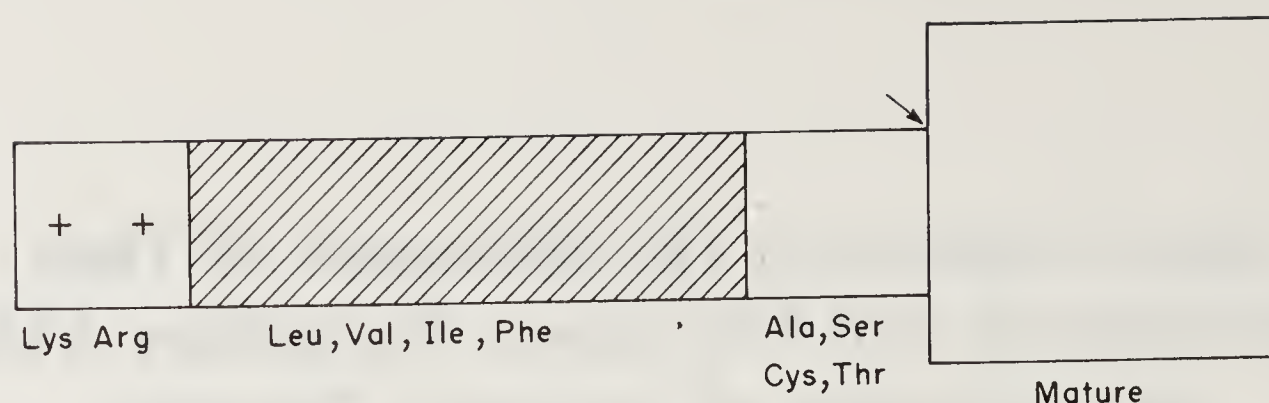


Figure 1. Pattern of occurrence of amino acids in secretory protein signal-sequences. Arrow indicates the point of cleavage from the mature portion. The hatched area corresponds to the hydrophobic segment.

Structural Aspects of Signal Sequences

Primary Structure of Signal Sequences

The primary structure of a large number of eukaryotic and prokaryotic signal sequences have been determined (15). A striking feature that is clearly discernible is the absence of any primary structure homology. However, a close examination of these sequences does reveal a pattern which is depicted in figure 1. All signal sequences without exception have a positively charged N-terminal region followed by a hydrophobic segment.

The importance of the charged and hydrophobic region of signal sequences with respect to function has been studied extensively by genetic techniques and *in vitro* reconstitution experiments (16, 17). The effects of various mutations on secretion are depicted in figure 2. Introduction of amino acids with charged side chains at the 9, 10 or 14th position in place of hydrophobic amino acids results in a signal sequence which is unable to initiate secretion in the case of *E. coli* alkaline phosphatase (18). Introduction of Gln, a polar amino acid, instead of Leu at position 8 and 14 also results in an inefficient signal sequence. However, a change involving a charged amino acid is much more lethal than a hydrophilic amino acid without a charge. A similar observation has been made with the signal sequence of *E. coli* λ receptor (19). The hydrophobic core of signal sequences has been replaced by a contiguous stretch of either Val, Leu or Ile residues. A minimum of 10 hydrophobic amino acids were found to be necessary for a functional signal sequence (20).

Conformation of Signal Sequences

The conformational preferences of signal sequences have been analysed by theoretical methods (21, 22). However, no consensus conformation emerges from these studies. In fact, theoretical analysis indicates that both helical and β -conformation are possible.

The first report of experimental studies on signal sequences was by Rosenblat *et al.* (23). They studied the conformation of the precursor region of parathyroid hormone in aqueous environment and in hexafluoroisopropanol. In aqueous environment, the signal sequence was found to adopt predominantly β -structure whereas in an apolar solvent an α -helical conformation was observed. Subsequently, studies on the

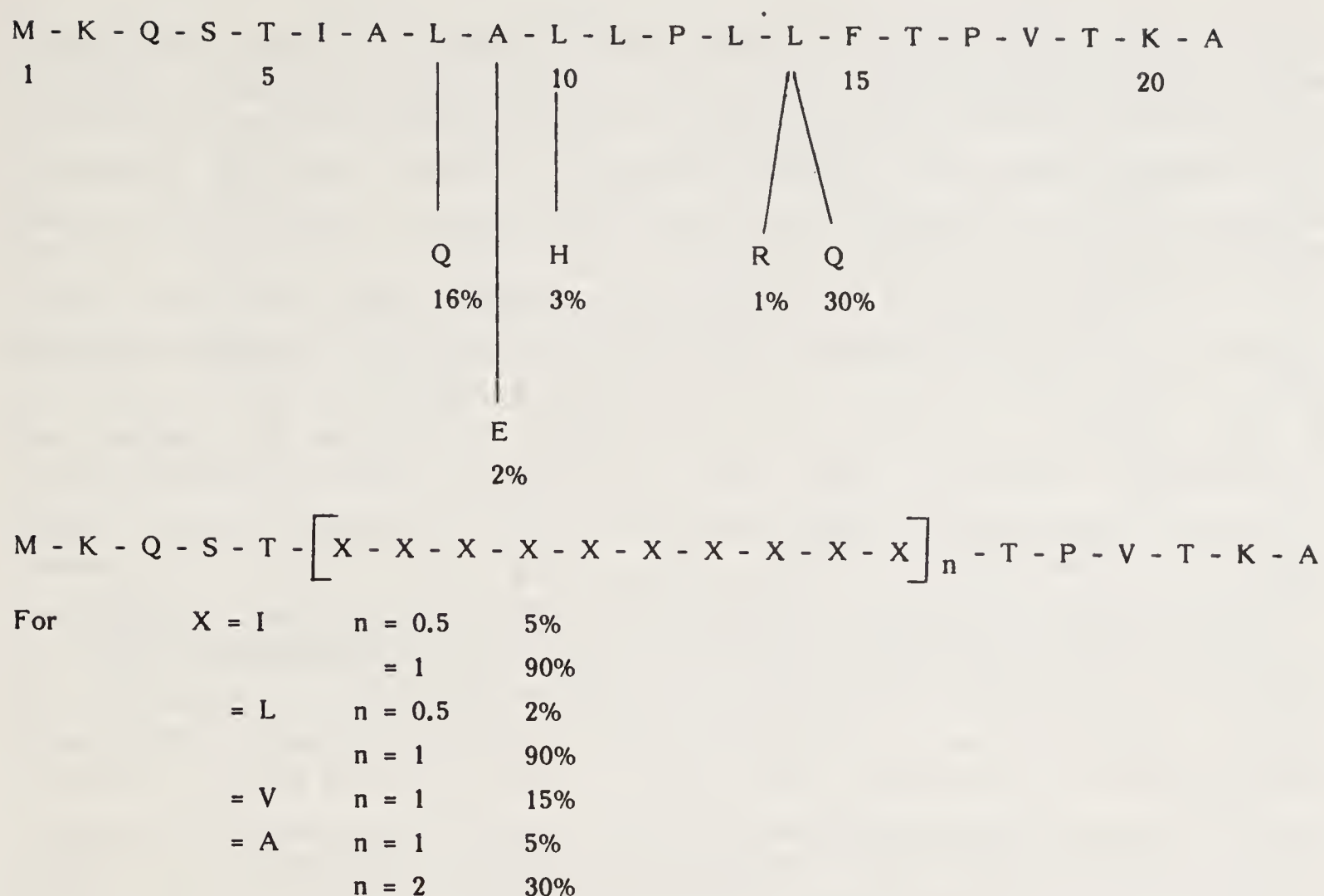


Figure 2. Effect of changes of amino acids in the signal-sequence of *E. coli* alkaline phosphatase. Percentages against amino acids correspond to amounts of mature protein detected with respect to wild-type signal sequence, when translocation is initiated by a signal-sequence with the indicated mutation. Data are from references 18 and 20.

conformation of several synthetic signal sequences in organic solvents and micelles of SDS have been reported (24–27). In organic solvents, particularly the apolar ones, it is evident that α -helical conformation predominates. The actual percentages should be viewed with caution as the standards for these estimations are proteins. Briggs and Gierasch have performed detailed studies on the signal sequence of the wild-type *E. coli* λ receptor, a deletion mutant which is defective in secretion and two revertants which restore export of λ -receptor (24). The wild type and mutant signals are shown in figure 3. The wild-type signal showed α -helical conformation whereas the “defective” mutant signal showed random conformation. The ‘revertants’ which were functional showed helical conformation. Theoretically predicted conformation of the wild-type, mutant and revertant λ receptor signal sequences and circular dichroism studies on peptides

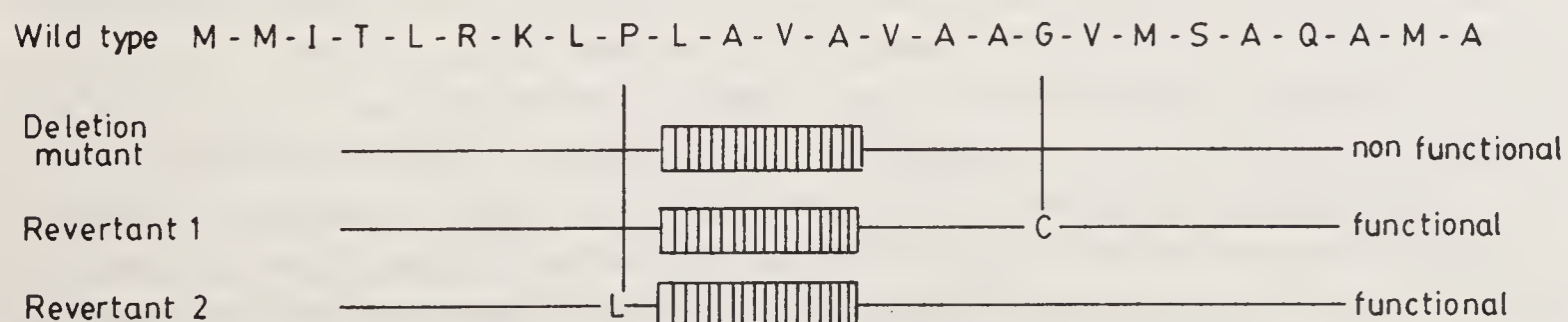


Figure 3. Wild-type, deletion mutant and revertant signal sequences of *E. coli* λ receptor (ref. 19).

corresponding to these sequences correlate well (28). Hence it was generalized that an α -helical conformation is probably the common recognition element in signal sequences. A close examination of the primary sequences of signals indicates the presence of Pro, a helix-breaking amino acid, at a central position, in several of them (15). In order to determine whether such signals have a high α -helical content the conformation of the alkaline phosphatase signal sequence, K-Q-S-T-I-A-L-A-L-L-P-L-L-F-T-P-V-T-K-A which has two Pro residues, one situated centrally at the 11th position and the other at the 16th position at the carboxy terminus, has been investigated in different environments (29, 30). In trifluoroethanol, methanol and aqueous mixtures of these solvents, the spectra were characteristic of peptides in α -helical and β -conformation. However, unlike some other signal sequences, predominantly β -structure was observed in SDS micelles. In order to investigate whether signal sequences can indeed adopt β -structures, particularly in hydrophobic environments, the conformation of wild-type *E. coli* lipoprotein, chicken lysozyme and mutant signals, derived from these were studied in media of increasing hydrophobicities. Media of different hydrophobicities were obtained with micelles of SDS, Brij and cholate (31). All wild-type signals showed a tendency to adopt β -structure in media of increasing hydrophobicities. Interestingly, mutant signals which have been shown to be non-functional adopted largely unordered structures in highly hydrophobic environment. β -structure has also been observed in the signal sequence of the gene product of *pho E* in hydrophobic micellar environment (32) as well as in *E. coli* λ receptor on the surface of the lipid bilayer (33).

The conformation of signal sequences is thus clearly solvent-dependent. In relatively polar environments both helical and β -conformations are observed, with helical conformation predominant in some and β -structure in others. Even some mutant non-functional signal sequences show ordered conformation in organic solvents and SDS micelles. Although organic solvents and SDS micelles offer a relatively apolar environment as compared to water, the hydrophobicity is not comparable to the hydrophobic interior of the lipid bilayer or even micelles of Brij or cholate. The marked preference of only wild-type and not non-functional mutant signal sequences to adopt the β -structure in highly hydrophobic environment strongly suggests that it is this structural motif that is recognized by components of the cells' export machinery.

It is reasonably clear that nascent secretory proteins bind to the signal recognition particle (SRP) as they emerge from ribosomes (34) and the complex is targeted to the ER membrane. The probable signal sequence binding subunit of SRP has been identified recently and this protein shares homology with a protein in *E. coli* (35). On the basis of theoretical analysis, a groove-like structure comprising helices with apolar surfaces has been proposed as the signal sequence-binding domain in these proteins (35). Based on this assumption, a possible mode of association of a signal sequence in a β -conformation in a hydrophobic pocket formed by two helices is shown in figure 4. In the case of mutant signals with a charged residue in the hydrophobic region, such an association would not be energetically favourable. In an helical conformation, the entire helix surface would be apolar in nature and hence association with two amphiphilic helices shown in figure 4 would not be favourable. Thus a β -structure would be more favourable than a helical structure for interaction with the proposed helical groove in the signal sequence binding domain in the SRP subunit.

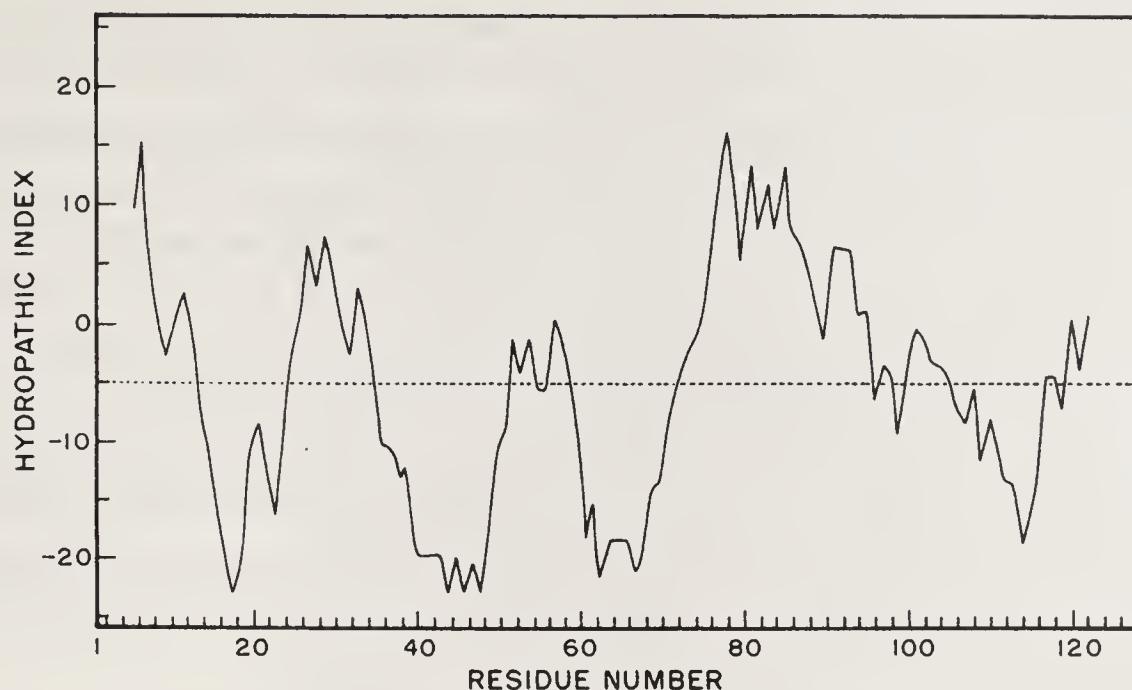


Figure 4. Hydropathic index plot of lysozyme. The lysozyme sequence was analysed by the method of Kyte and Doolittle (*J. Mol. Biol.* **157**, 505, 1982) averaging over a range of 9 amino acids. Hydrophobic domains are above the dotted line and hydrophilic domains are below.

Interaction of Signal Sequences with Membranes

The hydrophobic nature of signal sequences and the observation that thermodynamics would favour spontaneous association of signal sequences with lipids (36–38) lead to the proposal of models for protein translocation wherein the passage of proteins is envisaged to occur directly through the lipid bilayer (39, 40). The essence of these models is as follows: The signal sequences of nascent proteins bind to the head group of the phospholipids, a process facilitated by the positively charged amino terminus. Then the hydrophobic region enters the lipid bilayer followed by the mature portion of secretory proteins. Translocation is presumed to proceed directly through the lipid bilayer. During the process of secretion, the signal sequence is removed by signal peptidase. Engelman and Steitz have envisaged the translocation proceeding in a helical hairpin configuration (40). They have argued that the energy barrier that will have to be overcome for a hydrophilic peptide chain to traverse across the lipid bilayer is derived by the process of folding on the luminal side of the membrane.

In an effort to understand whether signal sequences indeed initiate translocation of proteins by directly partitioning into the lipid bilayer of membranes, several studies on the interaction of signal sequences with model membranes were undertaken (41–46). Fluorescence studies have indicated that signal sequences, particularly the hydrophobic portion, bind to model membranes and are associated with the hydrophobic core of the bilayer (41). It has further been shown that the association of signal peptides with model membranes results in extensive perturbation of the lipid bilayer so as to cause fusion of phosphatidyl serine vesicles and aggregation of phosphatidyl choline vesicles (43). The perturbation was such that the vesicles were rendered permeable to hydrophilic molecules like carboxyfluorescein. It has been argued that signal peptides even at very low concentrations have the ability to cause defects in the lipid bilayer. At increasing concentration of the peptide, the number of faults increase, resulting in fusion and aggregation. Recently, it has been shown that presence of very low concentrations of signal sequences can induce the formation of non-bilayer structures in model

membranes (46). Interestingly, even mutant signal sequences, with charged amino acids in the hydrophobic region were also found to have the ability to perturb model membranes (43, 45). The interaction of signal sequences with lipid monolayers has also been investigated. Signal sequences, differing widely in primary structures, were found to increase the surface pressure of monolayers indicating association with the acyl chains. In these experiments, mutant signals which could not initiate secretion were found to be less membrane-active as compared to the wild-type signals.

While it is reasonably clear that signal sequences spontaneously associate with membranes and can also cause perturbation at very low concentrations, photolabelling experiments suggest the presence of signal-sequence binding proteins in the ER membrane (47, 48) and *E. coli* inner membrane (49). However, it has not been clearly established that these proteins are indeed receptors for signal-sequences or whether they form channels to provide a pathway for protein export.

Signal sequences initiate translocation of proteins across membranes and are rapidly degraded even before the entire translocation is completed. Hence the presence of signal sequences is probably not necessary for *completion* of translocation across the membrane. The helical hairpin hypothesis which envisages translocation of proteins directly across the lipid bilayer may be a somewhat simplistic and unrealistic representation, since analysis of typical secretory proteins like lysozyme and RNase does not reveal the presence of any hydrophobic domain (figures 5 and 6). In fact the analysis strongly suggests a 'hydrophilic' path for translocation. Two recent observations, namely,

- (i) the presence of large aqueous channels in membrane vesicles derived from rough ER and plasma membranes of *E. coli* (50) and
- (ii) extractability of partially translocated nascent chains by 4 M urea (51), strongly support the view that peptide chain translocation occurs through a micro-environment that is accessible to aqueous reagents. However, there is no clear evidence that there are signal sequence receptors in the membrane which act as conduits on interaction with the signal sequence. We propose that signal

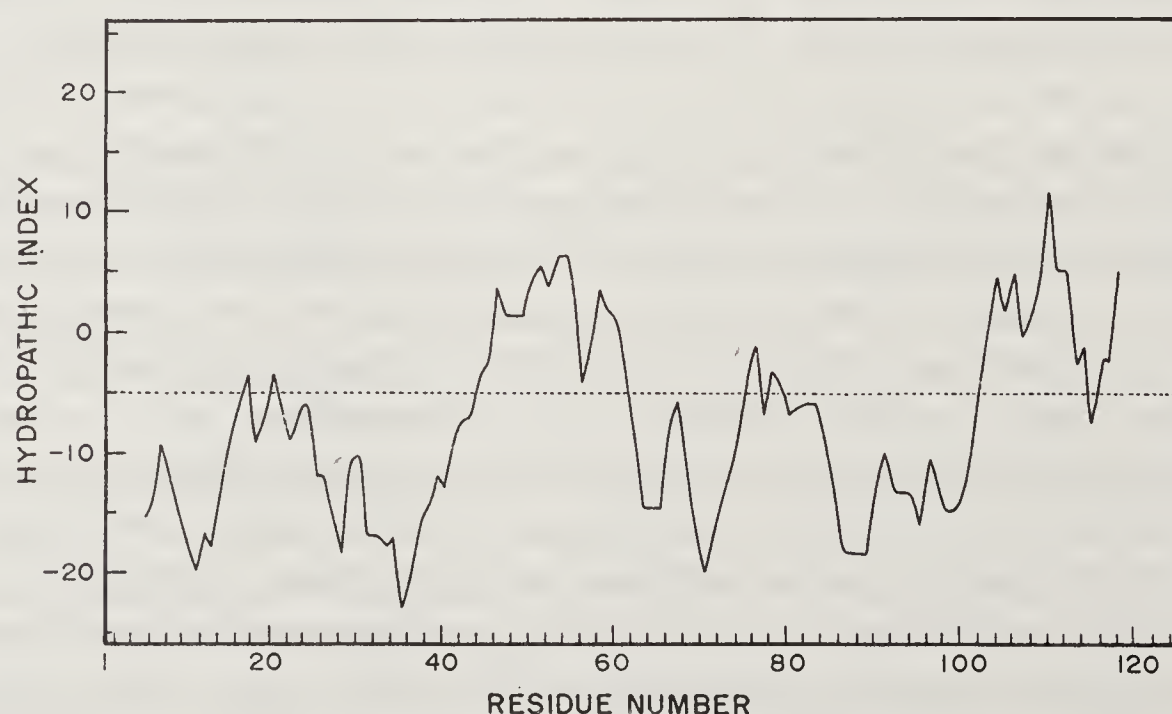


Figure 5. Hydropathic index plot of RNase A. The RNase A sequence was analysed by the method of Kyte and Doolittle (*J. Mol. Biol.* **157**, 505, 1982) averaging over a range of 9 amino acids. Hydrophobic domains are above the dotted line and hydrophilic domains are below.

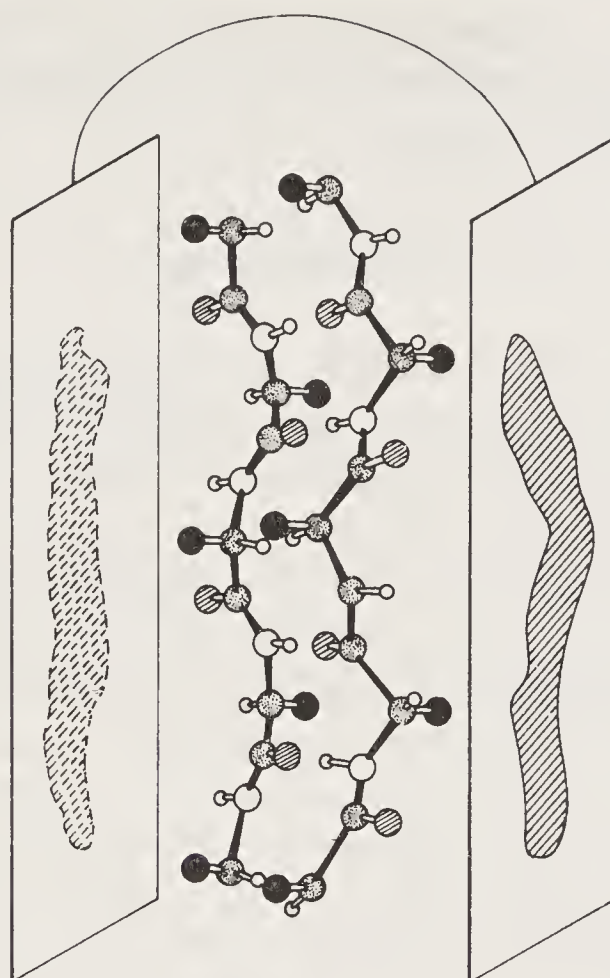


Figure 6. Schematic representation of the interaction of a portion of an anti-parallel β -sheet with hydrophobic surfaces of two helices. The two rectangles correspond to projections of an α -helix. The shaded portions correspond to apolar faces of amphipathic helices postulated by Bernstein *et al.* (ref. 35).

sequences spontaneously integrate into the lipid bilayer of translocation competent membranes. This association gives rise to 'faults' or 'defects' in the lipid bilayer which are responsible for generation of aqueous channels. Since mutant signals can also interact with membranes, it is likely that inability to initiate secretion stems from non-recognition by molecules like SRP based on conformational criteria discussed earlier.

There is considerable evidence in the literature that discrete contiguous peptide segments are a 'molecular address' for proteins destined for mitochondria chloroplasts (6), nucleus (7) and peroxisomes (5). Mitochondrial signals also occur at the amino terminus of premitochondrial or chloroplast proteins and are cleaved off after they have initiated passage into mitochondria or chloroplasts (6). Although mitochondrial signals differ from secretory protein signals in their amino acid composition and length, they have been shown to interact strongly with lipids and are presumed to initiate secretion by perturbing the mitochondrial membrane (52). Nuclear and peroxisomal proteins are not processed and the signals form a part of the protein. They are considerably shorter than mitochondrial and secretory protein signals (i.e. < 13 residues as compared to ~ 30 residues of a mitochondrial signal and ~ 25 residues of a secretory signal). They are also rich in basic amino acids. Photolabelling studies have suggested the presence of protein receptors for nuclear signals (53, 54). Preliminary studies from this laboratory indicate the absence of any specific receptor for peroxisomal signals. Studies based on recombinant DNA techniques have indicated that the tripeptide sequence Ser-Lys-Leu has the ability to direct even non-peroxisomal proteins into peroxisomes. However, using photolabelled Ser-Lys-Leu and longer peptides having this sequence i.e. Leu-Ile-Lys-Ala-Lys-Lys-Gly-Gly-

Lys-Ser-Lys-Leu we have not been able to detect any proteins specifically binding to these sequences in the peroxisomal membrane or intact peroxisomes (G. Lakshma Reddy and R. Nagaraj, unpublished results).

Conclusions

It is thus reasonably clear that a contiguous stretch of a peptide segment as small as 3 residues to about 40 residues can act as a molecular address and ensure correct targeting of proteins in cells. There is reasonable evidence that signal sequences are not only necessary but also sufficient to ensure proper targeting (34). Protein components of the cells' export machinery have also been identified and their postulated role in protein sorting is by either interacting with signal sequences or by maintaining secretory proteins in a 'translocation competent' conformation. Analysis of the results of conformational studies on secretory protein signal sequences and their interaction with membranes does, I think, provide reasonable answers to the questions raised in the Introduction. Firstly, the structural motif recognized in signal sequences by components of the cells' export machinery, particularly molecules like SRP, is a β -structure. Mutant signals that have a charged residue in the hydrophobic region cannot adopt β -structures when they encounter SRP and hence are not recognised and consequently cannot bring about targeting to the translocation competent membrane. Secondly, signal-sequences on encountering a membrane surface, spontaneously associate with lipids and cause 'defects' which are responsible in providing a hydrophilic microenvironment for translocation of the peptide chain across the membrane. However, we still do not know the structure of the 'hydrophilic' pathway in molecular terms that not only secretory proteins take but also proteins destined for peroxisomes, nucleus and mitochondria. Extensive research in various laboratories should hopefully provide answers in the years to come.

Acknowledgements

I am grateful to G. Lakshma Reddy and M. Joseph, who have contributed to investigation on signal sequences and whose work is quoted in this article. I am also grateful to Prof. D. Balasubramanian for discussion and critical evaluation of our circular dichroism results.

References

1. Silhavy, T. J., Benson, S. A. & Emr, S. D. (1983) *Microbiol. Rev.* **47**, 313-344.
2. Benson, S. A., Hall, M. N. & Silhavy, T. J. (1985) *Annu. Rev. Biochem.* **54**, 101-134.
3. Rapoport, T. A. (1985) *CRC Crit. Rev. Biochem.* **20**, 73-137.
4. Wickner, W. T. & Lodish, H. F. (1985) *Science* **230**, 400-407.
5. Osumi, T. & Fujiki, Y. (1990) *BioEssays* **12**, 217-222.
6. Hartl, F. U., Pfanner, N., Nicholson, D. W. & Neupert, W. (1989) *Biochim. Biophys. Acta* **988**, 1-45.
7. Gerace, L. & Burke, B. A. (1988) *Annu. Rev. Cell Biol.* **4**, 335-374.
8. Wolfe, P. B., Wickner, W. & Goodman, J. (1983) *J. Biol. Chem.* **258**, 12073-12080.
9. Evans, E. A., Gilmore, R. & Blobel, G. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 581-585.
10. Boehni, P., Gasser, S., Leaver, C. & Schatz, G. (1980) *The Organization and Expression of the Mitochondrial Genome*, eds. Kroon, A. M. & Csaccone (Elsevier, Amsterdam), pp. 423-433.

11. Walter, P., Gilmore, R. & Blobel, G. (1984) *Cell* **38**, 5–8.
12. Cabelli, R. J., Chen, L., Tai, P. C. & Oliver, D. B. (1988) *Cell* **55**, 683–692.
13. Watanabe, M. & Blobel, G. (1989) *Cell* **58**, 695–705.
14. Beckmann, R. P., Mizzen, L. A. & Welch, W. J. (1990) *Science* **248**, 850–854.
15. Watson, M. E. E. (1984) *Nucleic Acids Res.* **12**, 5145–5164.
16. Duffud, G. D., Lehnhardt, S. K., March, P. E. & Inouye, M. (1985) *Curr. Topics Membranes Transport* **24**, 65–104.
17. Bankaitis, V. A., Ryan, J. P., Rasmussen, B. A. & Banford, P. J. (1985) *Curr. Topics Membranes Transport* **24**, 105–150.
18. Michaelis, S., Hunt, J. F. & Beckwith, J. (1986) *J. Bacteriol.* **167**, 160–167.
19. Emr, S. D., Hedgpeth, J., Clement, J. M., Silhavy, T. J. & Hofnung, M. (1980) *Nature (London)* **285**, 82–85.
20. Chou, M. M. & Kendall, D. A. (1990) *J. Biol. Chem.* **265**, 2873–2880.
21. Austen, B. M. (1979) *FEBS Lett.* **103**, 308–313.
22. Garnier, T., Gaye, P., Mercier, J. C. & Robson, B. (1980) *Biochimie* **62**, 231–239.
23. Rosenblatt, M., Beaudette, N. V. & Fasman, E. D. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3983–3987.
24. Briggs, M. S. & Gierasch, L. M. (1984) *Biochemistry* **23**, 3111–3114.
25. Katakai, R. & Iizuka, Y. (1985) *J. Am. Chem. Soc.* **106**, 5715–5718.
26. Shinnar, R. & Kaiser, E. T. (1984) *J. Am. Chem. Soc.* **106**, 5006–5007.
27. Reddy, G. L. & Nagaraj, R. (1985) *Biochim. Biophys. Acta* **831**, 340–346.
28. Emr, S. D. & Silhavy, T. J. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4599–4603.
29. Reddy, G. L. & Nagaraj, R. (1986) *FEBS Lett.* **202**, 349–352.
30. Reddy, G. L. & Nagaraj, R. (1989) *J. Biol. Chem.* **264**, 16591–16597.
31. Shoba, J. & Balasubramanian, D. (1988) *Proc. Indian Acad. Sci. (Chem. Sci.)* **98**, 469–478.
32. Batenburg, A. M., Brasseur, R., Ruyschaert, J. M., Van Scharrenburg, G. J. M., Slotboom, A. J., Demel, R. A. & deKruiff, B. (1988) *J. Biol. Chem.* **263**, 4202–4207.
33. Briggs, M. S., Cornell, D. G., Dluhy, R. A. & Gierasch, L. M. (1986) *Science* **233**, 206–208.
34. Walter, P. & Lingappa, V. R. *Annu. Rev. Cell. Biol.* **2**, 499–516.
35. Bernstein, H. D., Poritz, M. A., Strull, K., Hoben, P. J., Brenner, S. & Walter, P. (1989) *Nature (London)* **340**, 482–486.
36. von Heijne, G. (1980) *Eur. J. Biochem.* **103**, 431–438.
37. von Heijne, G. (1981) *Eur. J. Biochem.* **116**, 419–422.
38. von Heijne, G. (1981) *Eur. J. Biochem.* **120**, 275–278.
39. DiRienzo, J. M., Nakamura, K. & Inouye, M. (1978) *Annu. Rev. Biochem.* **47**, 481–532.
40. Engelman, D. M. & Steitz, T. A. (1981) *Cell* **23**, 411–422.
41. Nagaraj, R. (1984) *FEBS Lett.* **165**, 79–82.
42. Briggs, M. S., Gierasch, L. M., Zlotnick, A., Lear, J. D. & DeGrado, W. F. (1985) *Science* **228**, 1096–1099.
43. Nagaraj, R., Joseph, M. & Reddy, G. L. (1987) *Biochim. Biophys. Acta* **903**, 465–472.
44. Batemberg, A. M., Deniel, R. A., Verkleij, A. J. & deKruiff (1988) *Biochemistry* **27**, 5678–5685.
45. McKnight, J. C., Briggs, M. S. & Gierasch, L. M. (1989) *J. Biol. Chem.* **264**, 17293–17297.
46. Killian, J. A., de Jong, A. M. Ph; Bijvelt, J., Verkleij, A. J. & deKruiff, B. (1990) *EMBO J.* **9**, 815–819.
47. Wiedmann, M., Kurzchalia, T. V., Hartmann, E. & Rapoport, T. A. (1987) *Nature (London)* **328**, 830–833.
48. Wiedmann, M., Kurzchalia, T. V., Bielka, H. & Rapoport, T. A. (1987) *J. Cell. Biol.* **104**, 201–208.
49. Bochkareva, E. S., Lissin, M. M. & Girshovich, A. S. (1988) *Nature (London)* **336**, 254–257.
50. Simon, S. M., Blobel, G. & Zimmerberg, J. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6176–6180.
51. Gilmore, R. & Blobel, G. (1985) *Cell* **42**, 497–505.
52. Roise, D., Horrath, S. J., Tomich, J. M., Richards, J. H. & Schatz, G. (1986) *EMBO J.* **5**, 1327–1334.
53. Adam, S. A., Lobl, T. J., Michell, M. A. & Gerace, L. (1989) *Nature (London)* **337**, 276–279.
54. Pandey, S. & Parnaik, V. K. (1991) *Biochem. Biophys. Acta* **1063**, 81–89.

Ionophores: Structure and Interaction in Relation to Transmembrane Ion-Transport

K. R. K. Easwaran

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

Introduction

One of the important functions of biological membranes is ion-transport which is crucial to all living cells and organisms (1). The transport of cations across biological membranes is usually facilitated by membrane integral proteins (2, 3). During the last few years, studies in the area of ion-transport have gained tremendous importance and several channel proteins, such as acetyl choline receptor, sodium and potassium channel, gap junction etc., have been identified as macromolecular pores (4, 5). Information processing by way of biological membranes is carried out on the basis of ion movements through various specific transmembrane protein channels. Ionic channels are fundamental elements in gating the flow of ions across cells and in maintaining the membrane potential. The translocation of ions through these channels are essentially modulated by discrete movements of the so-called gates which are controlled either by ligand binding, i.e. chemical transmitter (e.g. acetylcholine receptor), or by membrane electric field (e.g. sodium channel) or by both (e.g. calcium channel).

In spite of the developments of newer and advanced physicochemical and spectroscopic techniques, our present knowledge of the molecular mechanism of ion-transport facilitated by the membrane channel proteins is limited because of the difficulty in isolating; characterising and working out the structure and conformation of these relatively large-sized proteins. However, the progress achieved so far has been due to the discovery of a large number of specialized transport molecules (ionophores) (6–8). Ionophores are compounds produced by microorganisms which act by specifically increasing the ion-permeability selectively (9–12). The majority of them are formed by various species of *Streptomyces*, which are giant, spore forming, gram positive bacteria. The enniatins and alamethicin are from fungi (table 1). As can be seen in the table, ionophores can be divided as neutral carriers, which move as charged cation–ionophore complexes across the membrane, carboxylic carriers, which are believed to move across as neutral cation–anion pairs, and pores or transmembrane channels (models for channel proteins) through which ions can pass. The interesting observations are that they are found only in prokaryotes, but are effective on membranes of prokaryotic and eukaryotic cells. This leads to some interesting questions about their role in microorganisms. The medical and pharmacological applications of ionophores are limited as they lack specificity for

Table 1. Ion selectivity of ionophores (taken from ref. 7).

Class of Ionophores	Compound	Monovalent cation	K^+ / Na^+	Divalent cation
Neutral carriers	Valinomycin	$Rb > K > NH_4 > Na > Li$	10000	Weak complexation
	Nactins	$NH_4 > K > Rb > Na > Li$	210	
	Enniatin (B)	$NH_4 > K > Rb > Na$	37	
	Boromycin	$K > Na > Li$	1.7	Weak complexation
	Cyclic polyether XXXII	$Rb > K > NH_4$	21	
	Diamide ligands	$Na > K$ (weak complexation)	0.2	
Carboxylic carriers	Nigericin	$H; K > Rb > Na > Li$	45	$Ca > Mn > Sr > Mg$ $Ca > Mn > Sr > Mg$
	Monensin	$H; Na > K > Rb > Li$	0.1	
	Dianemycin	$H; Na > K > Rb > Li$	0.5	
	X-537	$H; K; Rb > Na > Li$		
	A23187	$H; Li > Na > K$	0.1	
Other carriers	Diphenylene Iodonium Proton conductors	H		
Channel formers (voltage independent)	Gramicidin A	$K > Rb > NH_4 > K > Na > Li$	1.8	
	Polyene antibiotics (Amphotericin B)			
	Alamethicin	$K > Rb > Na > Li$	1.7	Ca
(voltage dependent)	Monozomycin	$Na = K$		Ca

one type of membrane. However, they are found to be wonderful tools in delineating information on the molecular mechanism of ion-transport at the cellular or membrane level.

Ionophores can be essentially characterized as receptors, which form stable lipophilic complexes with ions such as K^+ , Na^+ , Ca^{2+} etc, and transport them across natural and artificial membranes. The processes of complexation and transport are often highly specific. There are essentially two mechanisms by which ionophores can lower the activation energy for movement for ions across membrane (figure 1). A carrier is a molecule that binds the ion and transports it as a carrier-ion complex through the membrane (e.g. valinomycin) whereas an ion channel is one which spans the entire membrane and contains in the interior a tunnel-like pathway for ions (e.g. gramicidin A).

Determination of the structure and conformation of the various ionophores and their complexes with cations of different size and charge and under a variety of conditions is very important for an understanding of the detailed mechanism of ion-transport across membranes mediated by these transporter molecules. During the last two decades, several groups around the world have worked on the structure,

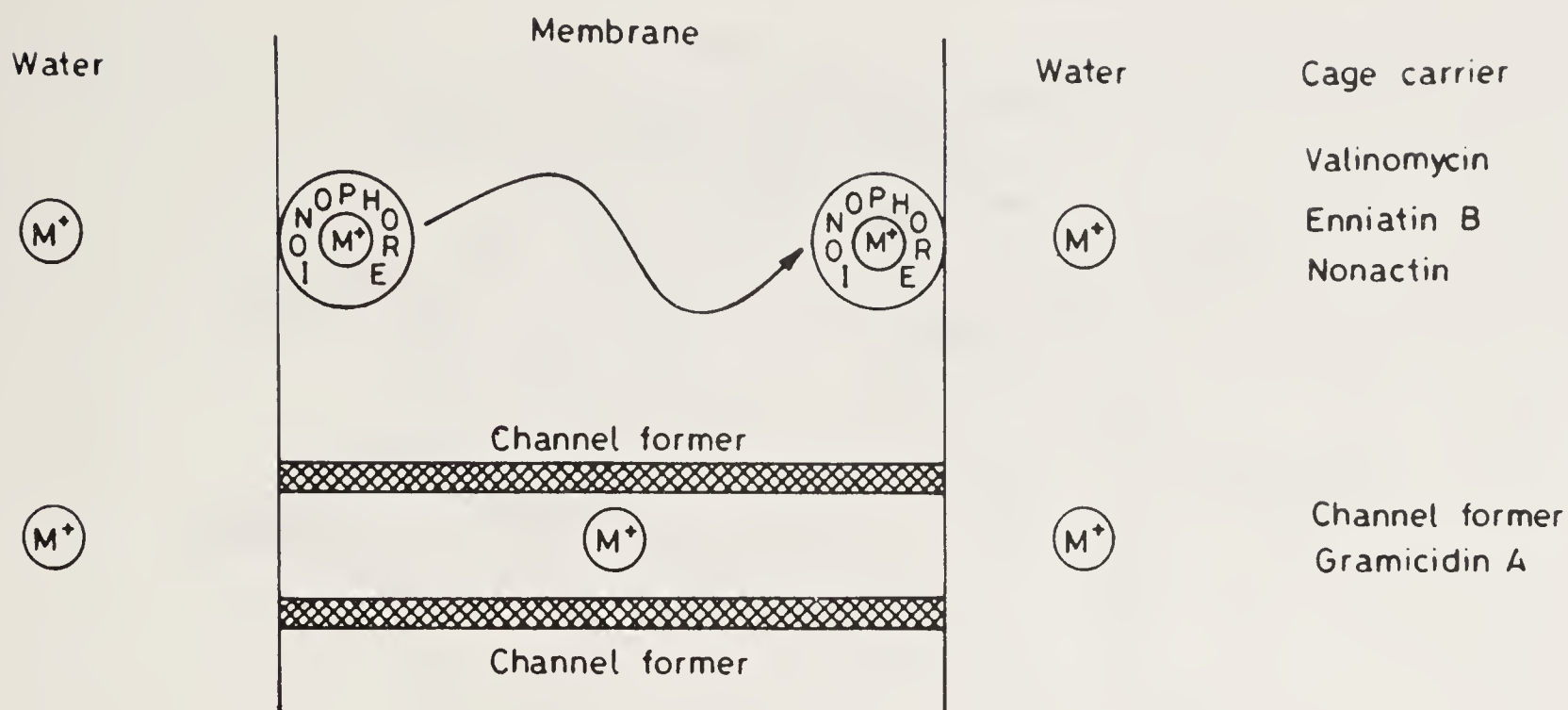


Figure 1. Mechanisms of cation transport.

conformation, interaction and dynamics of various natural ionophores, their synthetic analogs, and their cation complexes, and on the possible interaction of the ionophores with model and biological membranes using a variety of physio-chemical, spectroscopic and X-ray techniques (8, 12–19) with a view to formulate a molecular basis for the ion-capture and release by these transporters and for the transport mechanism in general. Our own group in Bangalore has contributed significantly in this area and this paper will focus on some of our data on two important ionophores, namely, valinomycin – a K^+ selective carrier, and calcimycin (A23187) – a divalent (Ca^{2+} and Mg^{2+}) selective carrier. Our results clearly show that the ion-selectivity of carrier ionophores depends not only on the nature of the ligands but also on the conformational state and properties of the whole molecule and also that conformation is very much dependent on the size and the charge of the ion and solvent and anion conditions.

1. Valinomycin (K^+ Ionophore)

Valinomycin (VM), a cyclic dodecadepsipeptide formed from alternate amino and hydroxy acids, consists of a trimer sequence, L-valine, D-hydroxy isovaleric acid, D-valine, L-lactic acid (figure 2), and is highly selective for K^+ ions (selectivity factor 10^4 over Na^+ ions). It is extensively used to specifically increase the K^+ ion permeability of biological membranes without affecting the permeabilities of other ions present and also to probe the nature of the K^+ -uptake system of a variety of cells and organells. The structure and conformation of free valinomycin and its K^+ complex both in solution and in the solid state have been worked out in detail using a variety of physico-chemical and spectroscopic techniques (8, 17, 20–24). It has been shown that valinomycin adopts different conformations depending on the polarity of the solvent (8). The three distinct conformations differ in the number of intramolecular hydrogen bonds (figure 3). It is interesting to note that in non-polar solvents the molecule adopts a well known bracelet conformation which is the most

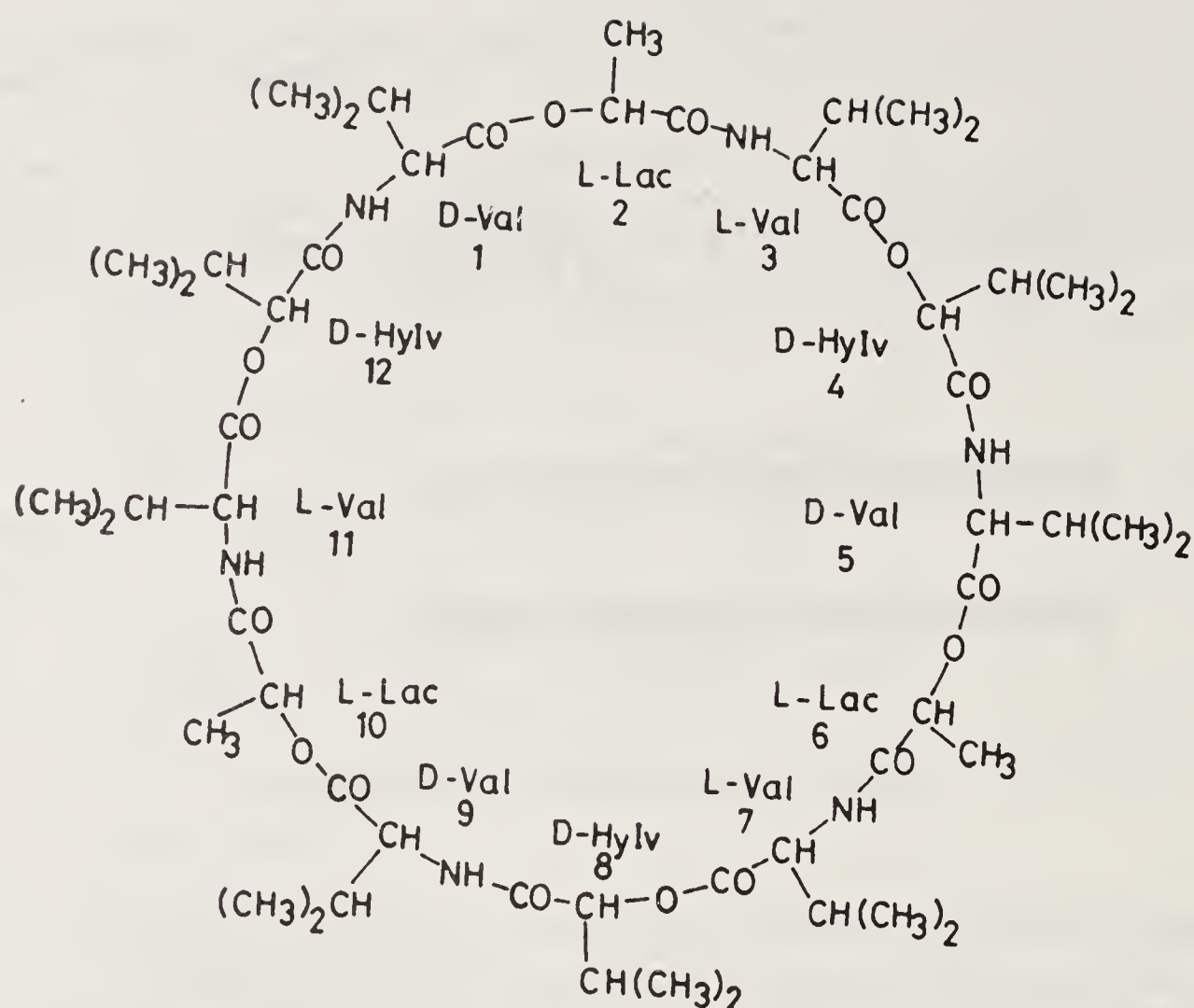


Figure 2. Chemical structure of valinomycin.

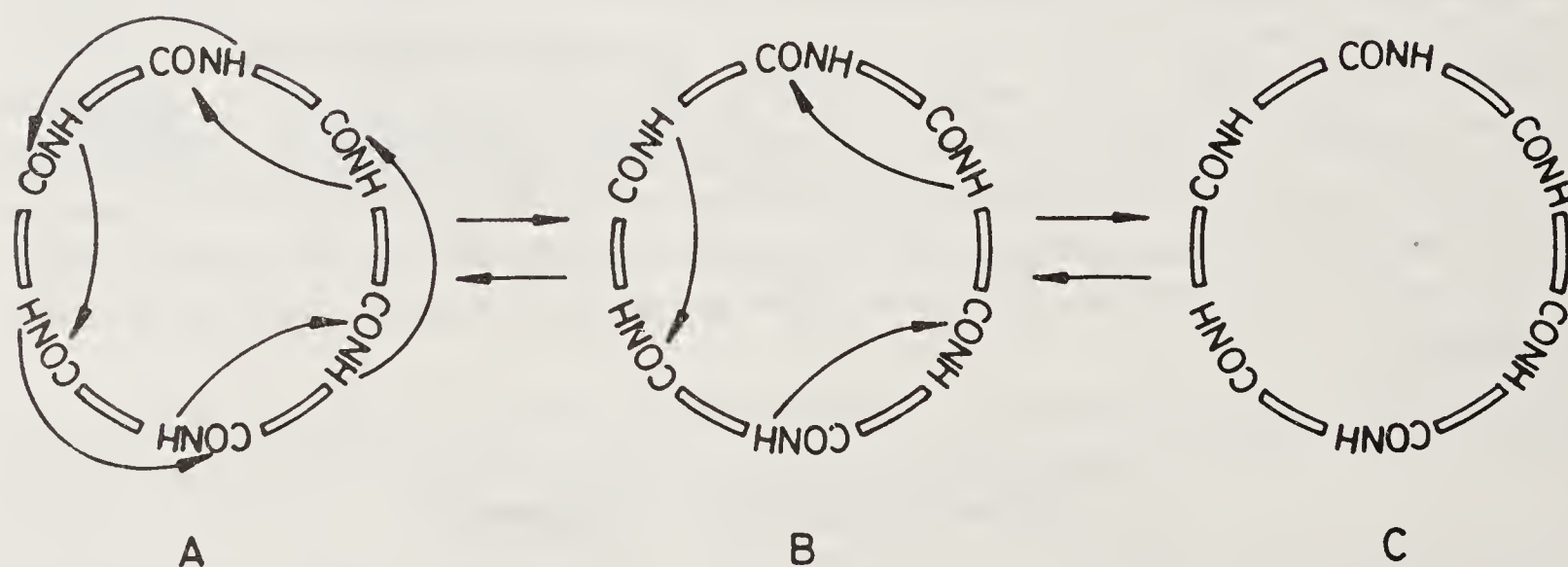


Figure 3. Conformations of valinomycin in (A) non-polar, (B) medium-polar, (C) polar solvents (taken from reference 8).

probable structure for the molecule in the interior of the membrane, whereas in solvents of medium polarity such as a 3:1 mixture of CCl_4 and DMSO, the molecule adopts an open or propeller conformation with only three intramolecular hydrogen bonds. The latter conformation is the most likely one at the membrane–aqueous interface. The solid state structure of free valinomycin crystallized in non-polar solvents is different from that proposed in solution (25). There are six intramolecular hydrogen bonds in the crystal structure, but the molecule is distorted with four $4 \rightarrow 1$ and two $5 \rightarrow 1$ intramolecular hydrogen bonds. However, the crystal structure of valinomycin crystallized from a medium-polar solvent, is very similar to that proposed

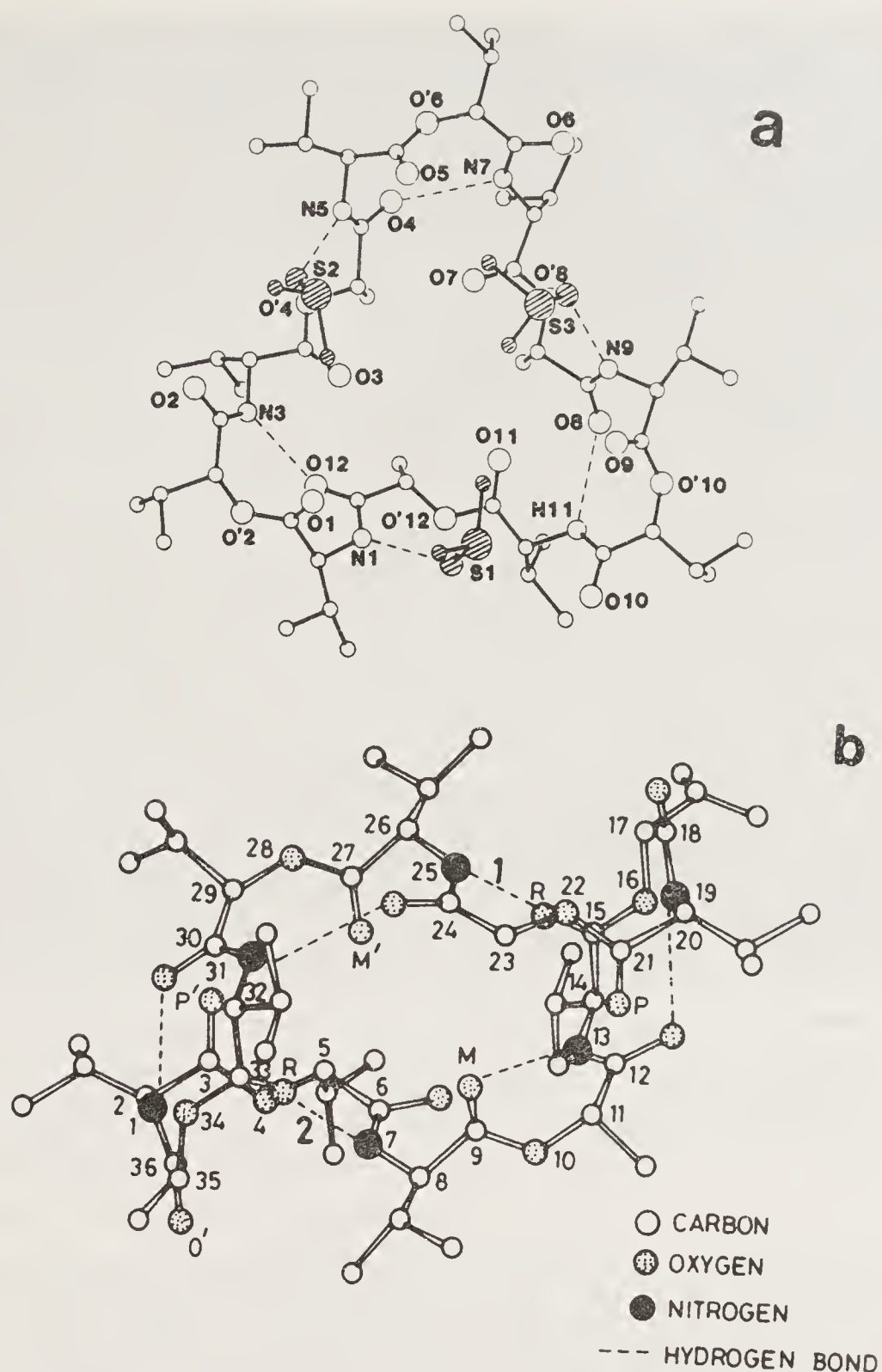


Figure 4. (a) Crystal structure of valinomycin crystallised in a medium-polar solvent, $(\text{CH}_3)_2\text{SO}$ (ref. 26). (b) Crystal structure of free valinomycin crystallised in a non-polar solvent (ref. 25).

in solution, namely a propeller type (26). Figure 4 shows a comparison of the crystal structure of free valinomycin in a non-polar solvent and that crystallized in a solvent of medium polarity.

The solution conformation and solid state structure of the VM-K^+ complex were very similar (17) with the molecule in the bracelet conformation having six intramolecular hydrogen bonds and the K^+ ion situated in the centre of the cavity liganding with six ester carbonyls. The molecular mechanism for the increase of K^+ transport across membranes by VM has been suggested as due to the diffusion of the VM-K^+ complex across the membrane interior after the molecule captures the K^+ ion at one interface and releases the K^+ ion at the other interface.

Detailed investigation carried out with other monovalent cations, Rb^+ , Cs^+ , Na^+ , and Li^+ indicates that the complexes of VM with Rb^+ and Cs^+ are very similar to the K^+ complex. The Na^+ complex differs from the well known VM-K^+ complex

in terms of the anion and solvent dependence and the nature of complexation (18). Detailed NMR studies of the VM-Li⁺ complex (20) shows that the Li⁺ forms an equimolar complex (different from the VM-K⁺ complex) with the cation complexed to the three ester carbonyls on the D-Val-L-Lac side of the bracket. A similar 1:1 peripheral complex with the cation complexed on the D-Val-L-Lac side of the bracelet structure has been observed for the VM-Na⁺ picrate complex studied by single crystal X-ray diffraction (27).

The conformations of VM-divalent cation complexes studied by a variety of techniques are quite different from those of VM-monovalent complexes (17). The conformations of VM-Ba⁺⁺ and Sr⁺⁺ complexes are very similar. Detailed CD and NMR studies carried out in complexes of VM with barium perchlorate and thiocyanate salts indicate that the ionophore forms a stable novel 1:2 (VM:barium) complex (22, 24). In solution the data indicate that the valinomycin forms complexes of stoichiometries 2:1, 1:1 and 1:2, with the most stable conformation being 1:2. This conformation can be described as an extended depsipeptide chain with no internal hydrogen bonds and wound in the form of an ellipse with two barium ions located at the foci. In this conformation the amide carbonyls are involved in the metal binding as opposed to the ester carbonyls in the VM-K⁺ complex. The single crystal X-ray structure analysis of the VM-barium perchlorate complex gives a flat open structure for this complex, with two barium ions per molecule similar to that proposed in solution (23). The barium ions are separated by 4.57 Å. Each barium ion is liganded to the three consecutive amide carbonyls, other atoms coordinating to the Ba⁺⁺ ions belonging to anion groups and solvent molecules. Figure 5 shows a comparison of the structure of VM-K⁺ and VM-Ba⁺⁺ complexes.

The significant differences in the structure of the two complexes, in spite of the similar ionic radii of K⁺ and Ba⁺⁺ ions, is possibly due to the differences in the hydrated radii of the two ions. Ba⁺⁺ with its larger hydration sphere is unable to penetrate into the cavity of the VM bracelet structure to form the ligand with the ester carbonyl oxygens due to the steric hindrance of the valyl side chains while the hydrated K⁺ ion has the optimum size to fit into the cavity and complex with the ester carbonyls. The conformational studies made with an analog of valinomycin, cyclo(Val-Gly-Gly-Pro)₃ in which the two D-residues of VM are replaced by glycine and L-Lac by L-Pro, thereby removing the steric hindrance, supported the above possibility. The proposed conformations for (Val-Gly-Gly-Pro)₃-Ba⁺⁺ and -K⁺ complexes were identical except for minor adjustments of the ester carbonyl group in the Ba⁺⁺ ion complex to compensate for the additional charge in the ion (28).

The solution conformation of VM-calcium complex arrived at from NMR and CD data showed that VM forms a 2:1 complex at lower salt concentration and a 2:1 and 1:1 peripheral complex co-existing at high salt concentration (17, 21).

2. A23187 (Ca²⁺ Ionophore)

Free A23187: A23187 (calcimycin), a divalent cation carboxylic ionophore, extensively used in studies of the role of the calcium ion in several cellular phenomena, is isolated from *Streptomyces chartreusensis* and has a structure consisting of a keto-pyrrole group, a spiroketal portion and a benzoxazole ring (figure 6). Extensive studies using

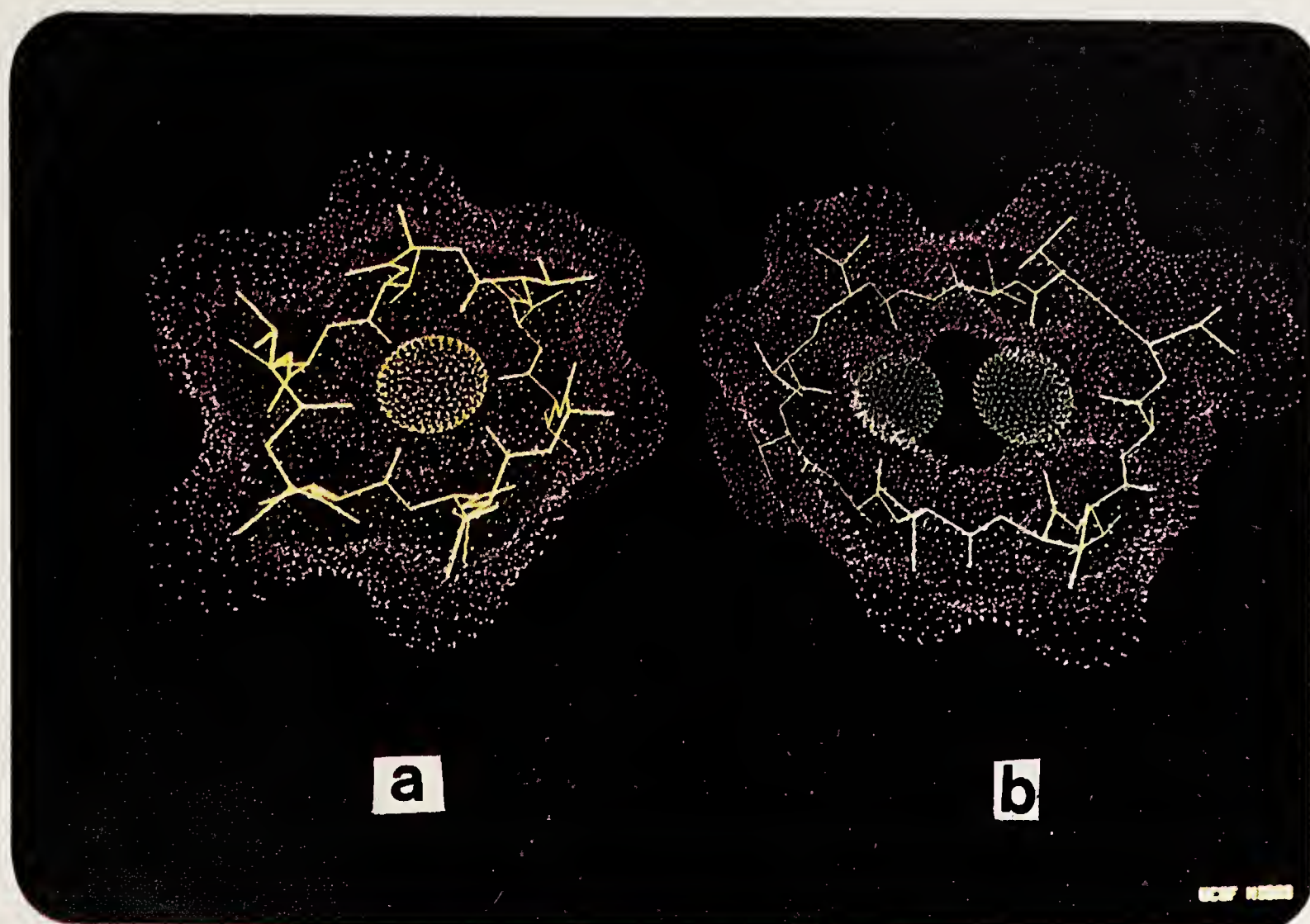


Figure 5. Crystal structure (a) valinomycin- K^+ complex (Neupert-Laves and Dobler, *Helv. Chim. Acta* **58**, 432, 1985); (b) valinomycin- Ba^{++} complex (ref. 23). (Figures generated using UCSF Midas software and LSM computer graphics facility at N. R. L., Washington, D.C. Solvent accessible molecular surfaces are indicated by dots).

UV, CD, fluorescence and NMR techniques on the free A23187 and its Ca^{++} and Mg^+ complexes have been carried out by various workers (29–35). Detailed NMR studies indicate that the free A23187 exists as a monomer in solution and also that its conformation is similar to that reported for its structure in the solid state (36). In the A23187- Mg^{++} and - Ca^{++} complexes it has been shown that two molecules of A23187 aggregate around the cation forming a 2:1 dimer complex with C_2 symmetry (33, 34). The crystal structure of the calcium complex of A23187 reported is very

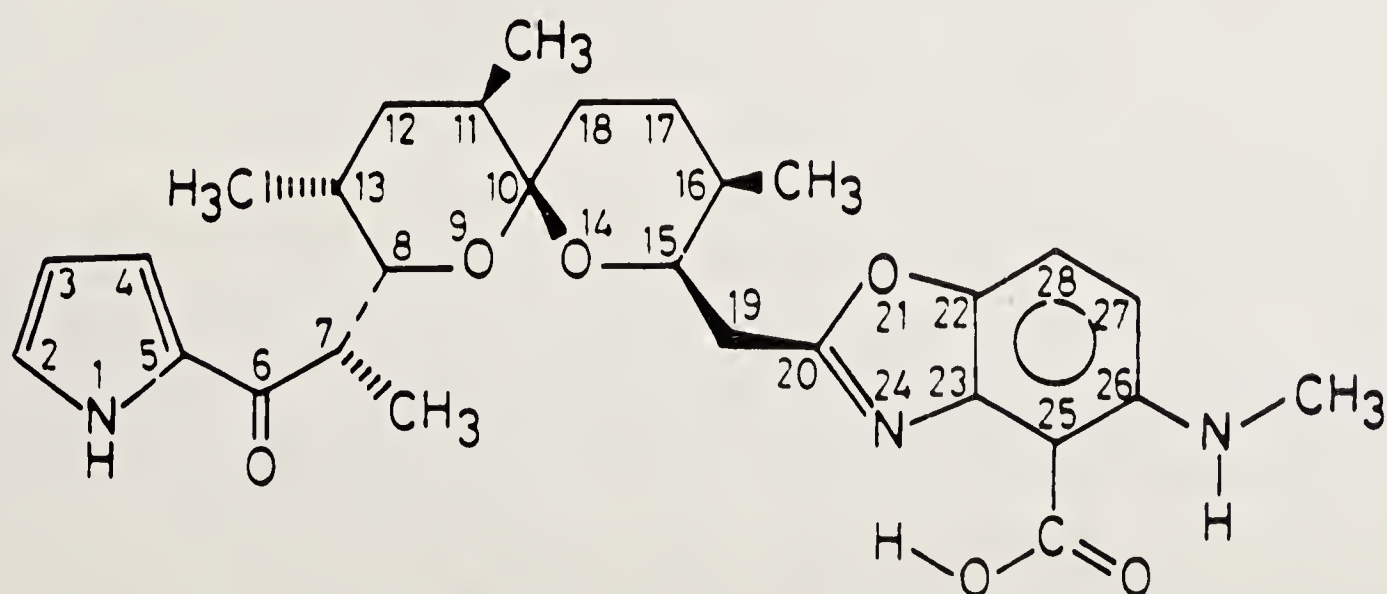


Figure 6. Chemical structure of A23187.

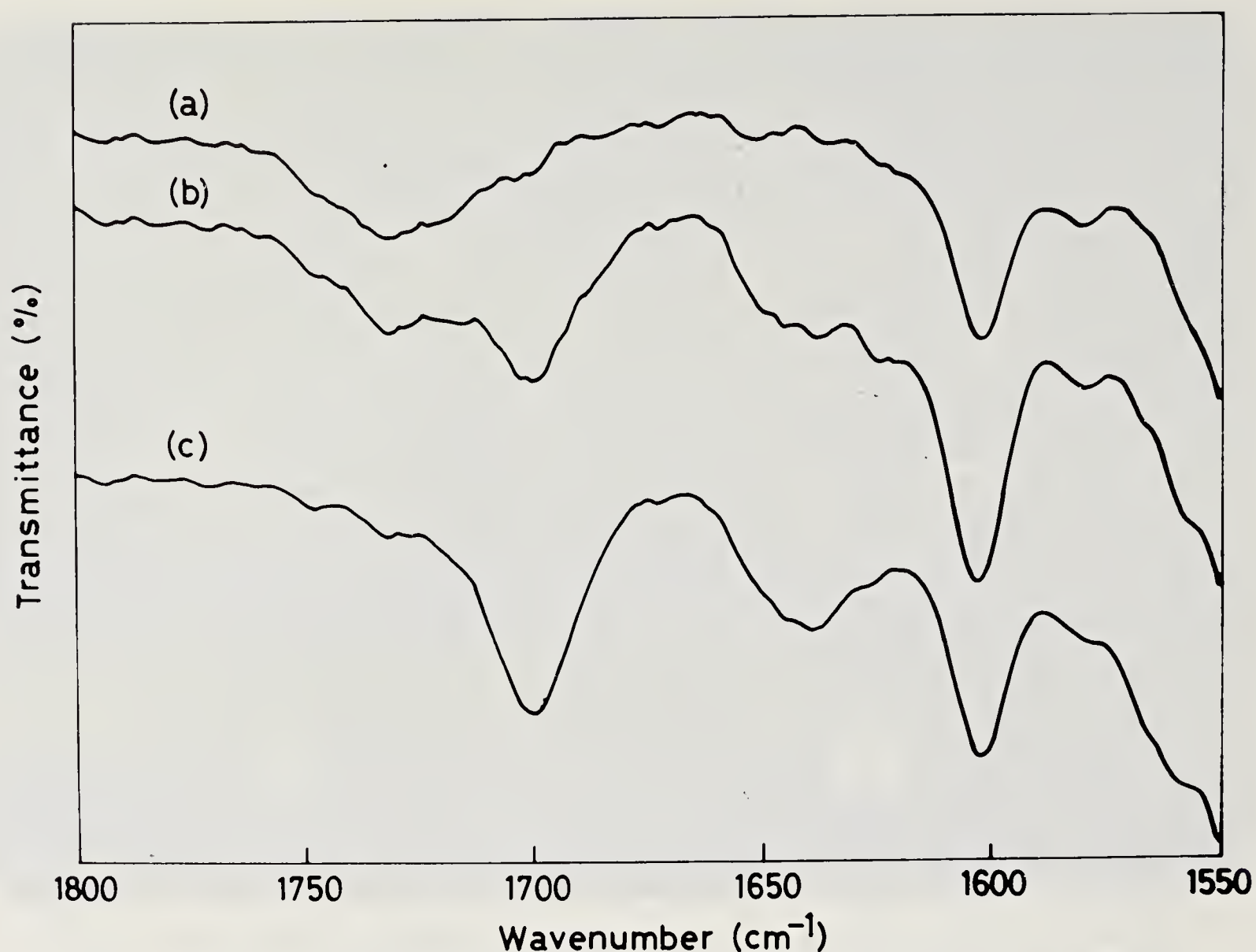


Figure 7. IR spectra of A23187 in CHCl_3 . Concentration of A23187 (a) 1×10^{-4} M, (b) 5×10^{-2} M and (c) 1×10^{-1} M.

similar to that obtained in solution by NMR forming a 2:1 dimer complex with the calcium ion bound to each ionophore molecule through a carbonyl oxygen, carboxyl oxygen and nitrogen of the benzoxazole ring system (37). However, our results on the structure and conformation of free A23187 as studied by IR, UV, CD, fluorescence, NMR and theoretical methods indicate that the free molecule also exists as a dimer in non-polar solvents and they further aggregate into a stacked dimeric pore structure in phospholipid vesicles. These results are presented below.

The UV spectrum of free A23187 in chloroform displays four resolvable bands at 380, 294, 280 and 245 nm (29). The peak intensity increases with increasing concentration and is saturated beyond 1.4×10^{-4} M indicating the possibility of association of the molecule into a dimer.

The IR spectra of A23187 in chloroform show significant changes with concentration (figure 7). With increase in the concentration of the ionophore, the band around 1730 cm^{-1} corresponding to the carboxyl $\text{C}=\text{O}$ decreases in intensity and disappears at high concentrations. However, the band around 1700 cm^{-1} corresponding to the ketopyrrole carbonyl group increases in intensity and is not shifted. An additional peak is observed at 1650 cm^{-1} which increases in intensity with concentration. The results show that as the concentration of the ionophore is increased the position of the ketopyrrole carbonyl peak is unchanged but the carboxyl $\text{C}=\text{O}$ stretching band is shifted from 1730 to 1650 cm^{-1} indicating that at higher concentration the

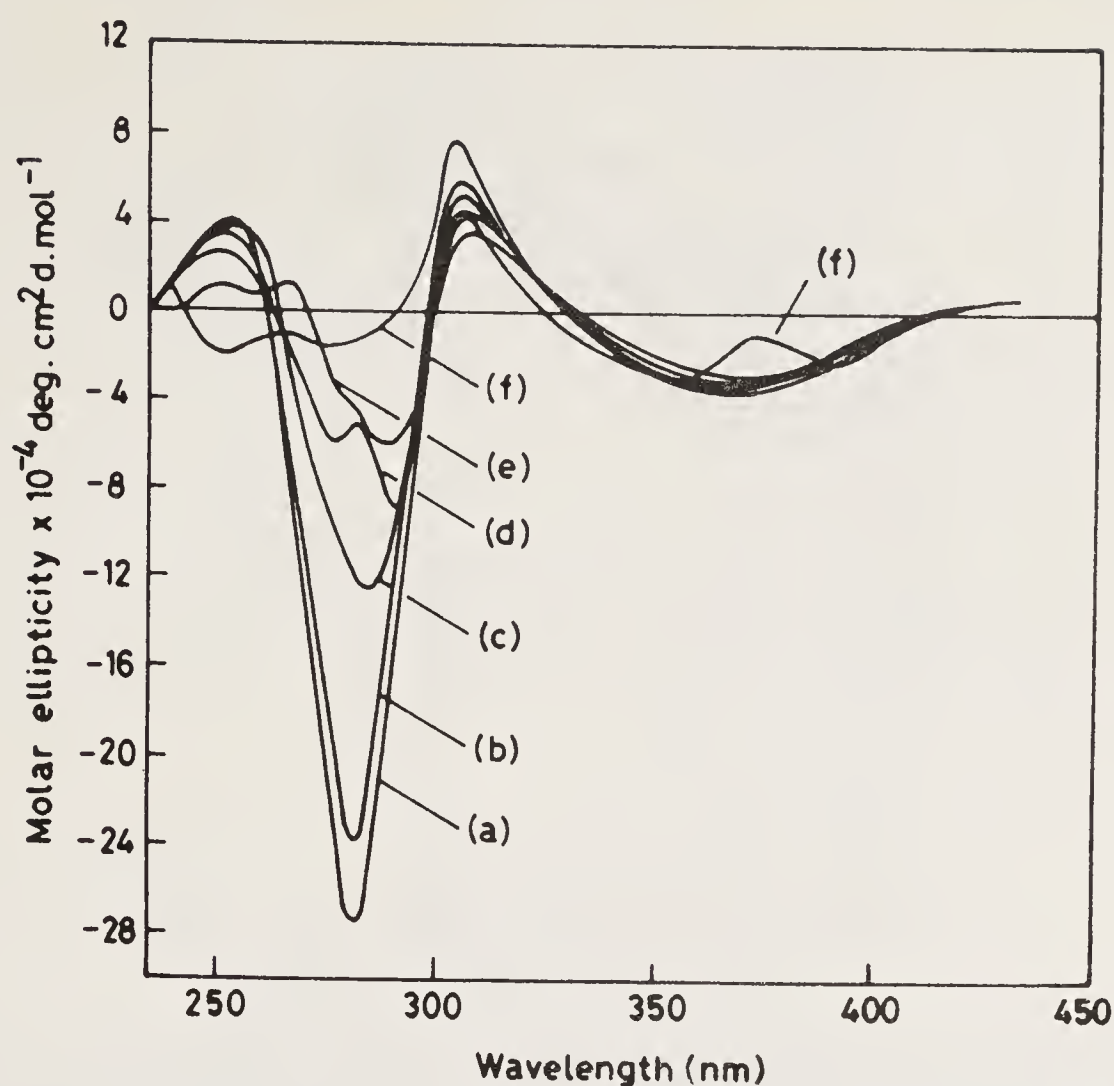


Figure 8. Time dependence of the CD spectra of A23187 (conc. = 1.05×10^{-4} M) (a) 0 min, (b) 60 min, (c) 3 h, (d) 5 h, (e) 16 h, and (f) 20 h.

carboxyl C=O is hydrogen bonded to the N-H group of the ketopyrrole group of an adjacent molecule resulting in a dimer (38).

The circular dichroism spectra of free A23187 in CHCl_3 showed concentration and time dependence (31). At high concentration (1.0×10^{-2} M) a large positive band is observed around 305 nm which diminishes in intensity with decrease in concentration. However, the band around 280 nm is very small in intensity at high concentration as compared to the very large negative ellipticity at lower concentration (e.g. at 1×10^{-3} M concentration of A23187). But the interesting observation is that the CD spectra of the ionophore at lower concentration (1×10^{-3} M) show remarkable time dependence (figure 8) with spectra stabilizing after several hours. Final CD spectra are very similar to those observed at high concentrations. No time dependence is observed in the CD spectra taken at high concentrations, i.e. greater than 10^{-2} M. The above results clearly point to the fact that the molecule has a tendency to aggregate as a dimer in non-polar solvents such as chloroform. It is also observed that the time taken for the molecule to form a dimeric structure depends on the concentration of the ionophore, the lower the concentration, the more time it takes to aggregate as a dimer.

The fluorescence spectrum of A23187 in chloroform also shows significant quenching of the fluorescence intensity with time. This again supports the evidence of dimer formation. The final spectrum is very similar to that of the spectra for the A23187- Ca^{2+} complex (figure 9).

The one-dimensional proton NMR studies on A23187 in CDCl_3 reported earlier suggest that the molecule exists as a monomer and the conformation in solution is

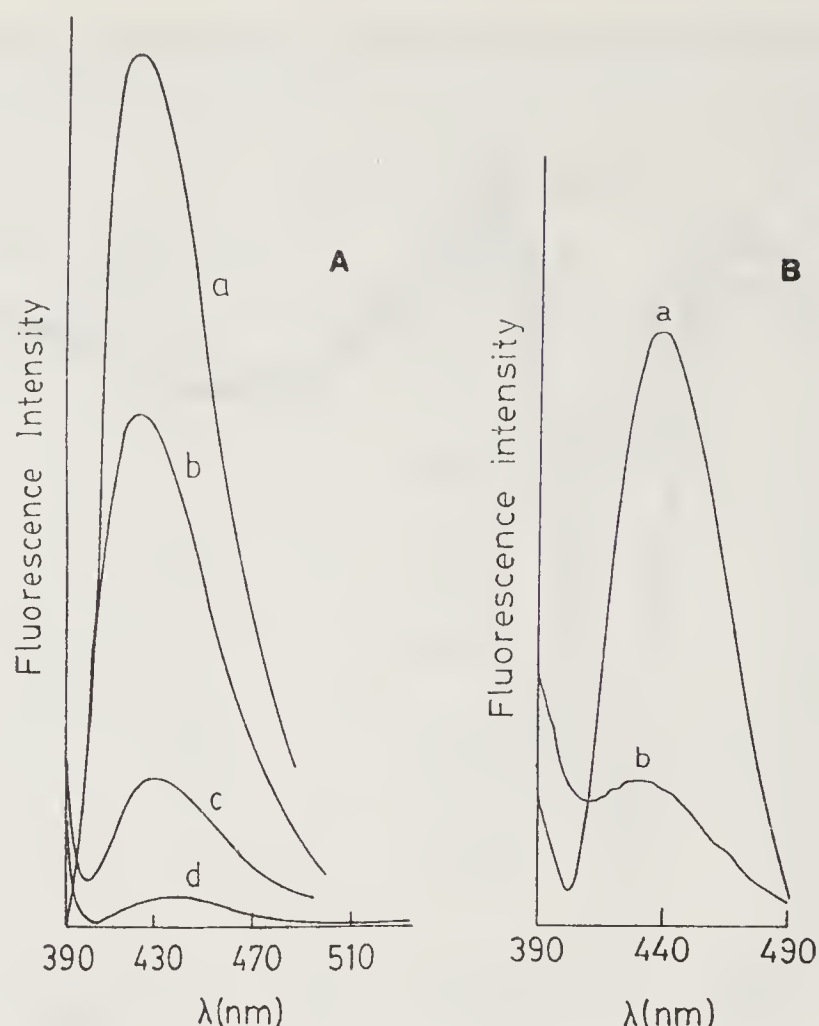


Figure 9. Fluorescence spectra. (A) A23187 in CHCl₃ as a function of time: (a) within minutes of the preparation of the sample, (b) 3 h, (c) 8 h and (d) after 24 h. (B) (a) A23187 in CHCl₃, (b) A23187-Ca²⁺ complex in CHCl₃ (1:2). Concentration of ionophore is 8×10^{-7} M.

similar to that reported from X-ray structure (36). However, our results on the solution conformation of A23187 in CDCl₃ is in disagreement with that proposed by earlier workers. We have carefully assigned the various resonances using the 2-dimensional COSY technique (38). The one-dimensional NOE data (table 2) show unusual NOE between certain protons, which cannot be explained as arising from a single molecule. Careful analysis of the spectra and the NOE data, and model building show that the data agrees well with the dimeric structure for the molecule.

Additional support to the dimeric structure for A23187 comes from circular dichroism data, which show that this ionophore forms complexes with leucine and other amino acids at low ionophore concentrations ($< 10^{-4}$ M) and the absence of calcium ion binding in the presence of amino acids at low concentrations of

Table 2. Proton NOE data^a for A23187 in CDCl₃.

Proton irradiated	Proton for which NOE observed
26 (NH ⁻ Me)	25 (COOH), 11, 12
11, 12	26 (NH ⁻ Me), 25 (COOH)
28	27, 11, 12
15	8

^aSee figure 6 for the numbering scheme.

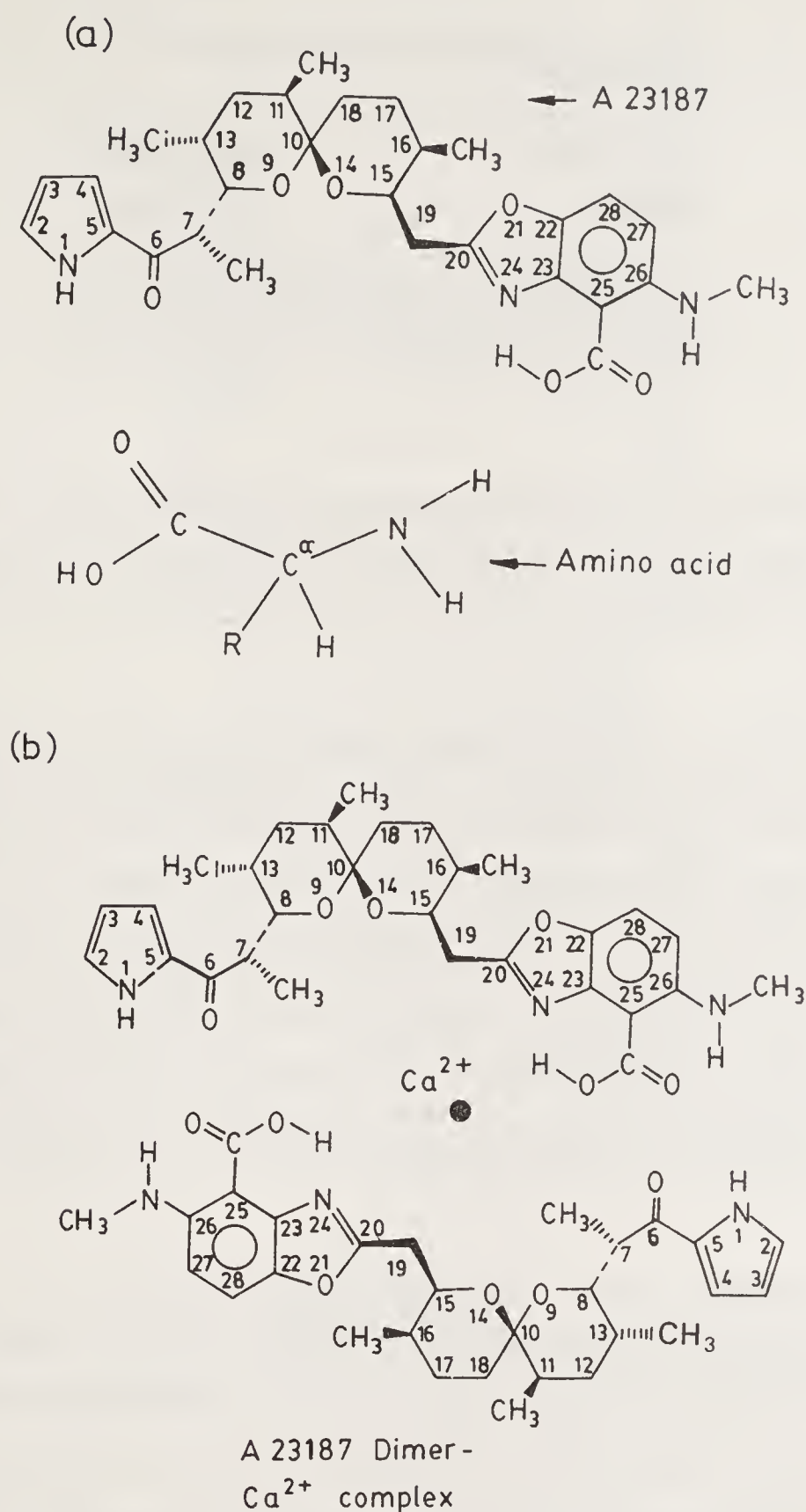


Figure 10. Proposed (a) A23187–amino acid complex – at low ionophore concentration. (b) A23187 dimer– Ca^{2+} complex – at high ionophore concentration.

A23187. At high concentrations of A23187 ($> 10^{-4}$ M) amino acids do not bind to the ionophore whereas there is a strong binding to the calcium ion. These results point to the fact that at lower concentrations the ionophore is a monomer and amino acids easily bind to the molecule thereby preventing dimerization and binding of Ca^{2+} or Mg^{2+} . At high concentrations, it easily dimerises and does not form a complex with the amino acid but forms an ideal structure for binding to Ca^{2+} or Mg^{2+} ion (39) (figure 10). The possibility of amino acids interacting with A23187 only at low concentrations was reported earlier on the basis of the solvent phase extraction procedure (40).

We have also determined the minimized energy generated by CHARMM (using the quanta software of the Polygen Corporation and silicon graphics work station) using the crystal structure data of free A23187. Results show that the dimeric structure

Table 3. Minimized energy generated by CHARMM*.

Free A23187	A23187 dimer
6.29 kcal	– 14.75 kcal

*Co-ordinates obtained from crystal structure of M. O. Chaney *et al.* (1974) *J. Am. Chem. Soc.* **96**, 1932.

formed with hydrogen bonds between ketopyrrole NH and C=O of the COOH group of benzoxazole ring has much lower energy than the monomer (table 3).

Based on all the data presented above, it is clear that the most stable form of free A23187 is dimer.

A23187–Ca²⁺ complex: Detailed NMR studies carried out on A23187 divalent (Ca²⁺ and Mg²⁺) show significant changes in the chemical shift and coupling constants as compared to the free spectra. These can be explained as due to the changes induced by the divalent cation co-ordinating to the carboxyl oxygen, a carbonyl oxygen and nitrogen of the benzoxazole ring system of each ionophore in the dimeric structure. It has also been shown by earlier workers (33, 34) that in the A23187–Mg²⁺ complex two monovalent units aggregate around the divalent cation forming a 2:1 dimer complex with a cavity of approximately 3.5 Å diameter and exhibiting a C₂ symmetry axis. The crystal structure of the calcium ion complex of A23187 reported is very similar to that obtained in solution forming a dimer 2:1 ionophore: cation complex (37).

A23187 in phospholipid vesicle: The fluorescence spectrum of A23187 in DPPC vesicle exhibits decrease in fluorescence intensity with time (38). The fluorescence polarization value measured for A23187 in DPPC (0.43) is an order of magnitude larger than that observed in CHCl₃. These results point to the possibility of further aggregation of the dimer molecule in a lipid environment. The CD spectrum of A23187 in DPPC ULV's also show remarkable time dependence (31). The spectra taken within minutes of sonication are similar to that observed at high concentrations of ionophore in non-polar solvents such as chloroform except for an additional induced peak around 345 nm. However, the spectra show continuous changes with time and stabilize after 40 min. The significant change of the CD band at 300 nm due to the ketopyrrole C=O group with time is a clear indication of the involvement of this group in the formation of an aggregated structure in the phospholipid vesicle possibly forming a hydrogen bond with the NH of the benzoxazole group of the adjacent dimeric structure.

Our CD and fluorescence results clearly show the possibility of an aggregated dimeric structure of A23187 in phospholipid vesicles with the NH's of the benzoxazole ring of one dimeric structure hydrogen-bonded to the C=O's of the ketopyrrole group of the adjacent dimer. Model building shows that about 5 dimers of A23187 would be sufficient to span a bilayer thickness forming a stacked pore (figure 11). The results presented here indicate the possibility of transmembrane calcium transport by a diffusive carrier. Our recent studies on the A23187 induced bilayer conductance

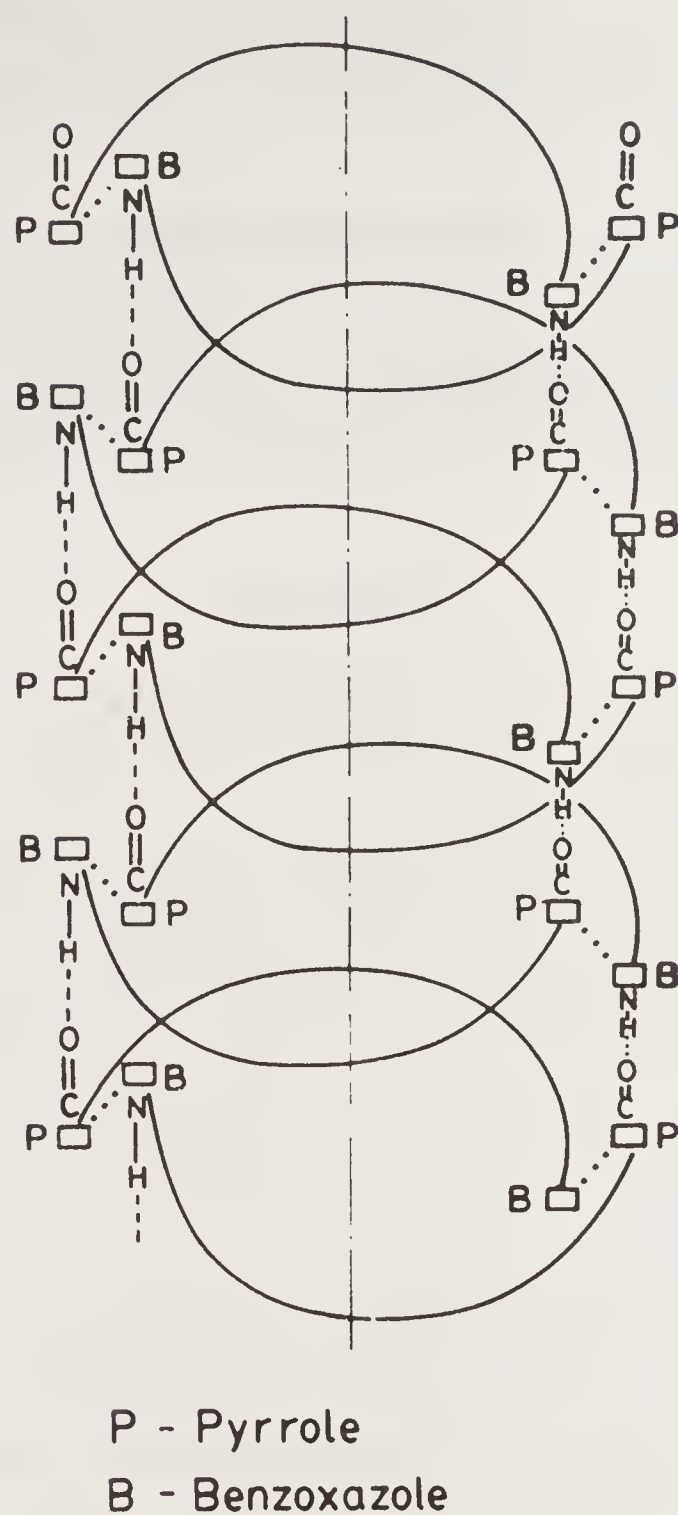


Figure 11. Proposed model of stacked dimeric pore structure of A23187 in DPPC vesicle.

measurements give an unusual calcium ion conducting profile more like that of a channel rather than that of a diffusive carrier.

Summary

Studies aimed at the understanding of the ionophoric abilities of carriers reveal their propensity to form ion-specific stable ionophore-cation complexes. The structures and conformations of the ionophores and their cation complexes are very much dependent on the nature of the cation (their size, charge and hydration properties) and the solvent polarity. Precise knowledge of the conformational changes of the ionophore molecule under different conditions is needed to understand the detailed mechanism of ion capture and release as has been illustrated taking valinomycin and A23187 as examples. Our results on the calcium ionophore, A23187, clearly show that the molecule exists as a dimer in non-polar solvents and aggregates into a stacked

dimeric pore in phospholipid vesicles leading to the possibility of calcium transport across membranes by a pore rather than by a diffusive carrier mechanism.

Acknowledgements

The untiring efforts of my associates, Late Dr. S. Devarajan, Drs. C. K. Viswanath, M. B. Sankaram and Mr. S. V. Balasubramanian who have contributed significantly to the studies reported in this article are gratefully acknowledged.

References

1. Jain, M. K. (1988) *Introduction to Biological Membranes* (John Wiley & Sons, New York).
2. Giebisch, G., Tosteson, D. C. & Ussing, H. H. (1980) in *Membrane Transport in Biology* (Springer-Verlag, Berlin) Vol. 3.
3. Scarpa, A. & Azzone, G. F. (1970) *Eur. J. Biochem.* **12**, 328–335.
4. Hille, B. (1984) *Ionic Channels of Excitable Membranes*, (Sinauer Associates, Sunderland).
5. Keeling, D. & Benham, C. (eds.) (1989) *Ion Transport* (Academic Press, London).
6. Pressman, B. C. (1976) *Annu. Rev. Biochem.* **45**, 501–530.
7. Bakker, E. P. (1979) in *Antibiotics*, ed. Hahn, F. E. (Springer-Verlag, Berlin) Vol. 5, pp. 67–97.
8. Ovchinnikov, Yu. A., Ivanov, V. T. & Shkrob, A. M. (1974) in *Membrane Active Complexones* (Elsevier, Amsterdam).
9. Pressman, B. C., Harris, E. J., Jagger, W. S. & Johnson, J. H., (1967) *Proc. Natl. Acad. Sci., U.S.A.* **58**, 1949–1956.
10. Andreoli, T. E., Tieffenberg, M. & Tosteson, D. C. (1967) *J. Gen. Physiol.* **50**, 2527–2545.
11. Mueller, P. & Rudin, D. O. (1967) *Biochem. Biophys. Res. Commun.* **26**, 398–404.
12. Szabo, G., Eisenman, G., Ciani, S. (1969) *J. Membrane Biol.* **1**, 346–382.
13. Ivanov, V. T. (1975) *Ann. N. Y. Acad. Sci.* **264**, 221–247.
14. Grell, E., Funck, T. & Eggers, F. (1975) in *Membranes. A Series Advances*, ed. Eisenman, G. (Marcel Dekker, New York) Vol. 3, Chap. 1.
15. Ovchinnikov, Yu. A., & Ivanov, V. T. (1975) *Tetrahedron* **31**, 2177–2209.
16. Prestegard, J. H. & Chan, S. I. (1969) *Biochemistry* **8**, 3921–3927.
17. Easwaran, K. R. K. (1985) in *Metal Ions in Biological Systems*, ed. Sigel, H. (Marcel Dekker, New York) Vol. 19, 109–137.
18. Easwaran, K. R. K. (1987) in *Ion Transport through Membranes*, eds. Yagi, K. & Pullman, B. (Academic Press, Tokyo) pp. 17–39.
19. Grell, E., Funch, Th. & Santer, H. P. (1973) *Eur. J. Biochem.* **34**, 415–424.
20. Sankaram, M. B. & Easwaran, K. R. K. (1982) *Biopolymers* **21**, 1557–1567.
21. Vishwanath, C. K. Easwaran, K. R. K. (1982) *Biochemistry* **21**, 2612–2621.
22. Devarajan, S. & Easwaran, K. R. K. (1981) *Biopolymers* **20**, 891–899.
23. Devarajan, S. & Easwaran, K. R. K. (1984) *Int. J. Pept. Protein Res.* **23**, 324–333.
24. Devarajan, S., Nair, C. M., Vijayan, M. & Easwaran, K. R. K. (1980) *Nature (London)* **286**, 640–641.
25. Smith, G. D., Duax, W. L., Lang, D. A., deTitta, G. T., Edmonds, J. W., Rohrer, D. C. & Weeks, C. W. (1975) *J. Am. Chem. Soc.* **97**, 7242–7247.
26. Karle, I. & Flippen-Anderson, J. L. (1988) *J. Am. Chem. Soc.* 3253–3257.
27. Steinraf, L. K., Hamilton, J. A. & Sabesan, M. N. (1982) *J. Am. Chem. Soc.* **104**, 4085–4091.
28. Easwaran, K. R. K. Pease, L. G. & Blout, E. R. (1979) *Biochemistry* **18**, 61–67.
29. Pfeiffer, D. R., Reed R. W. & Lardy, H. A. (1974) *Biochemistry* **13**, 4007–4013.
30. Puskin, J. S., Visities, M. I. & Wene, M. T. (1981) *Arch. Biochem. Biophys.* **206**, 164–172.
31. Balasubramanian, S. V. & Easwaran, K. R. K. (1989) *Biochem. Biophys. Res. Commun.* **158**, 891–897.
32. Divakar, S. & Easwaran, K. R. K. (1987) *Biophys. Chem.* **27**, 139–147.

33. Anteunis, M. J. O. (1977) *Bioorg. Chem.* **6**, 1–11.
34. Deber, M. & Pfeiffer, D. R. (1976) *Biochemistry* **15**, 132–140.
35. Caswell, A. H. & Pressman, B. C. (1972) *Biochem. Biophys. Res. Commun.* **49**, 292–298.
36. Chaney, M. O., Demaro, P. V., Jones, N. D. & Occolowitz, J. C. (1974) *J. Am. Chem. Soc.* **96**, 1932–1933.
37. Smith, G. D. & Duax, W. L. (1976) *J. Am. Chem. Soc.* **98**, 1578–1580.
38. Balasubramanian, S. V. & Easwaran, K. R. K. (1990) in *Water and Ions in Biomolecular Systems* (Birkhauser Verlag, Basel) pp. 221–228.
39. Balasubramanian, S. V. & Easwaran, K. R. K. (unpublished results).
40. Hovi, T., Williams, S. C. & Allison, A. C. (1975) *Nature (London)* **256**, 70–72.

Structure and Assembly of Viruses

M. R. N. Murthy¹, C. N. Hiremath¹ and H. S. Savithri²

¹Molecular Biophysics Unit and ²Department of Biochemistry,
Indian Institute of Science, Bangalore 560012, India

Introduction

A detailed understanding of the relationship between molecular structure and function is one of the fundamental aims of modern biology. The structure, organization and assembly of viruses have served as excellent model systems towards this goal. As obligate parasites, viruses specifically infect and proliferate in their respective host cells. The smallest infectious unit of a virus is called a virion. The genetic material (RNA or DNA) in virions is encapsidated in a protein shell called capsid. The function of the capsid is primarily to protect the genome in the harsh extracellular environment. Outside the living cell, the smaller viruses possess precise molecular structures which enable X-ray diffraction studies on their crystalline preparations. Information on the structure and assembly of viruses has been obtained by physical methods such as sedimentation behaviour, electron microscopy, neutron and X-ray low angle scattering, X-ray crystallography, and solution properties such as protection of nucleic acid by capsid protein aggregates, stability of virus particles as a function of temperature and ionic strength, in the presence or absence of divalent cations, detergents and other chemicals. Many of the commonly occurring plant viruses are easily purified in large quantities and pose limited environmental hazards. Hence, a large body of early information was gathered on plant viruses.

Extensive structural and assembly studies have been carried out with small ssRNA containing viruses and certain complex bacteriophages. The present review is, however, restricted to the studies on ssRNA viruses.

Architecture of Small RNA Viruses

The limited genetic information on viral nucleic acids is insufficient to code for an asymmetric protein large enough to encapsidate the genome. Hence, all viral capsids are constructed from a large number of either identical or only a few distinct classes of protein subunits. Small RNA viruses generally have genomes of size 1.5 to 4 million daltons, sufficient to code for the coat protein and a few others required for its multiplication in the host cell. The coat proteins are generally 20–40 kd in size. A large number of them are assembled to form the capsid of size (200–350 Å diameter) sufficient to package the genome (1, 2). Occurrence of a large number of identical protein subunits leads to a high degree of symmetry in the capsid structure (3). Energetically

favourable structures are likely to be obtained by repeating the same set of optimal bonding contacts between identical protein subunits. Unsymmetrical structures have different patterns of non-covalent bonding between chemically identical protein units in different regions of the capsid leading to ambiguities in assembly and increase of the free energy of the assembled complex. Hence, all small ss-RNA viruses are found to be based on either helical or icosahedral symmetry.

Helical arrangement of protein subunits is a simple way of repeating identical contacts between chemically indistinguishable protein subunits. A helical arrangement is characterized by the number of subunits (n) per turn and the pitch or axial rise per turn (h). The number of subunits per turn need not be integral as was realized in the case of alpha helical peptides by Pauling (4). Except for the asymmetry at the ends, the *vulgare* strain of tobacco mosaic virus (TMV) forms a perfectly helical structure with $16\frac{1}{3}$ subunits and 23 Å axial rise per turn (5). In contrast, in the *dahlemense* strain of TMV, the subunits are constrained to have nearly undeformable side-to-side contacts. However, the contact between the subunits in successive turns (axial contacts) of the helix exhibit periodic fluctuations leading to a smooth curvature for the rod-shaped particle (6). However, deformation in the axial contacts are small and hence the subunits remain in quasi-equivalent environments. Flexible, sinuous filamentous virus particles are probably constructed from similar distortions in bonding patterns.

Identical inter-subunit bonding is also achieved in closed symmetric shells. Closed capsids with symmetries of the three platonic solids, tetrahedron, octahedron and icosahedron accommodate 12, 24, or 60 subunits, respectively, in identical environments. Electron microscopic and X-ray diffraction studies have established the icosahedral symmetry in the capsids of several isometric viruses. A regular icosahedron is constructed from 20 triangular faces and has 532 symmetry (figure 1). The universal occurrence of icosahedral symmetry in the capsids of all the viruses examined so far

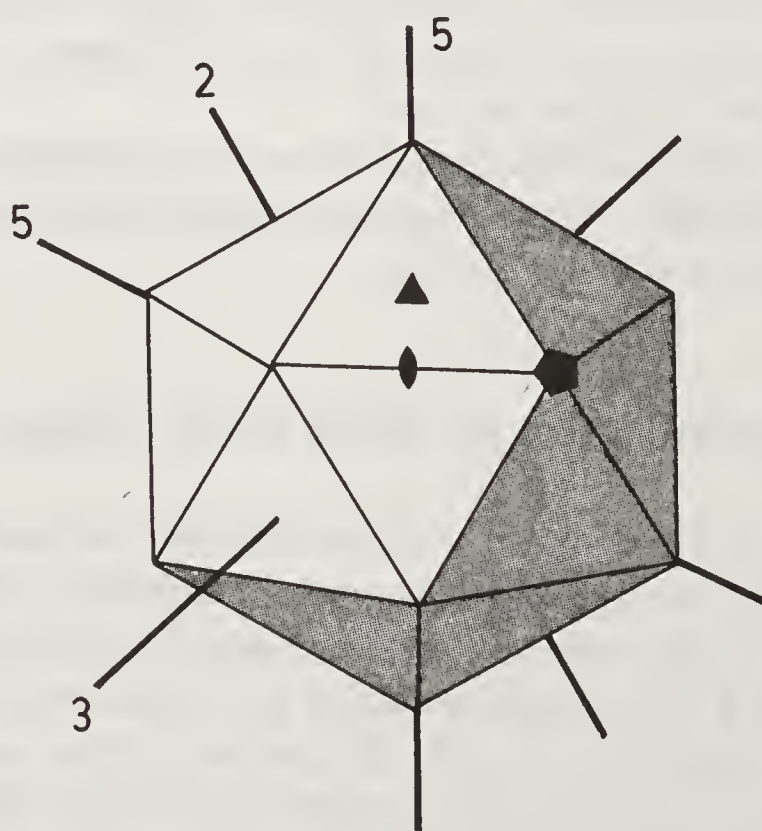


Figure 1. Icosahedron. All known isometric viruses possess icosahedral symmetry. An icosahedron is constructed using 20 equilateral triangles which form a closed symmetric solid characterised by 12 five-fold, 20 three-fold and 30 two-fold axes.

suggests some general principles underlying the assembly of stable isometric particles. The number of protein units that can be organized with identical surroundings on the surface of a sphere is highest (3 per triangular facet, a total of 60) when the symmetry is icosahedral. Use of this symmetry leads to the largest interior volume per protein unit available for packaging the nucleic acid.

Quasi-Equivalence

Icosahedral symmetry can accommodate only 60 protein subunits in identical environments. However, biochemical studies revealed the presence of more than 60 subunits in several viral capsids. These subunits cannot be accommodated with identical inter-subunit contacts. Also, several spherical viruses reveal 32 distinct morphological units in their negatively stained electron micrographs. There is no simple way of organizing 60 icosahedrally related protein units into 32 clusters. In order to account for these observations, Caspar and Klug (7) proposed the theory of *quasi-equivalence*. According to this theory, it is not necessary to organize protein subunits in identical environments for achieving stable, minimum energy structures. Identical subunits need to occupy only quasi-equivalent positions. It was shown that certain multiples of 60 subunits could be organized to have nearly equivalent, if not strictly identical, bonding interactions if certain design principles are adhered to. The quasi-equivalence of chemically identical subunits was assumed to result from slight deformation of bonding interactions in non-symmetry related environments. A systematic procedure for the construction of the various icosahedrally symmetric polyhedra (icosa-deltahedra), accommodating quasi-equivalently related subunits was worked out (7). The procedure consists of converting hexameric clusters at 12 specific chosen lattice points into pentameric clusters.

A regular icosahedron (figure 1) has 20 equilateral triangular faces. A deltahedron is a closed shell of equilateral triangles. An icosa-deltahedron is a deltahedron with icosahedral symmetry. Different classes of icosa-deltahedron can be constructed from hexagonal nets of equilateral triangles by removing a triangle from twelve selected lattice points and folding the net so as to connect the cut ends. An icosa-deltahedron has $20T$ facets, where T is a triangulation number, whose value is restricted by $T = Pf^2$, where f is any integer, and $P = h^2 + kh + k^2$, for all pairs of integers h and k having no common factors. Protein subunits can be accommodated with quasi-equivalent environments only when the capsid conforms to one of the icosa-deltahedral classes. Icosahedral lattices corresponding to $T = 1, 3$, and 4 have mirror symmetry. $T = 7$ and several other higher triangulations have two distinct enantiomorphic possibilities. Each structural unit in the shell can be visualized as one-third of a triangular facet of the icosa-deltahedron, so that there are $20T \times 3 = 60T$ quasi-equivalent subunits in a shell. These protein subunits might cluster in different ways as dimers, trimers or pentamers and hexamers to give characteristic appearance to the virus particles. By considering the possible ways of clustering structural units, the theory provides methods for predicting gross features of viral morphology. For example, 12 clusters of pentameric structural units and $10(T - 1)$ clusters of hexameric structural units is one of the often observed arrangements for capsid structure (figure 2). Detailed analysis of the inter-subunit contacts in southern bean mosaic virus (SBMV) which is based on a $T = 3$ lattice shows that the protein subunits do not have quasi-equivalent bonding. The

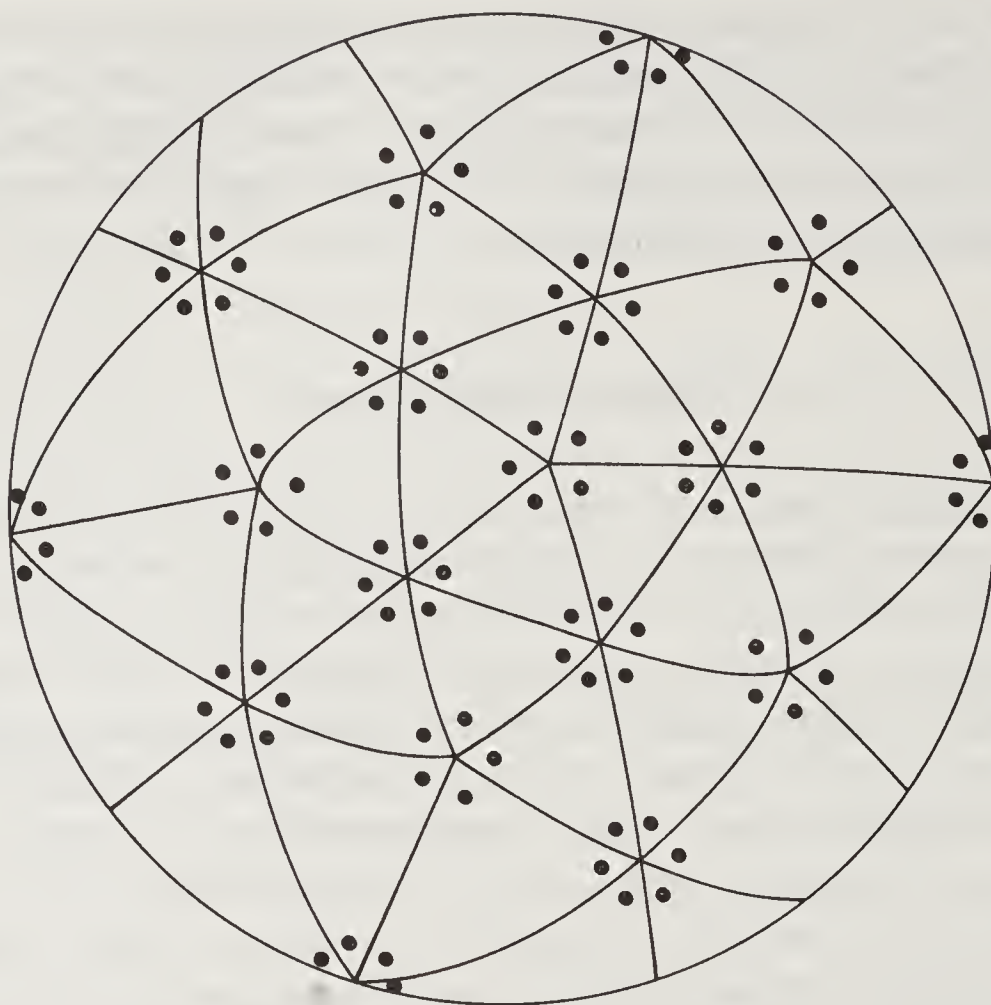


Figure 2. Schematic diagram illustrating one of the possible ways of protein subunit clustering on the surface of an isometric virus particle based on $T=3$ icosahedral lattice. Each black dot represents a protein subunit.

implications of these observations for the error-free, self-assembly of viruses are presented in the section on virus assembly.

Structural Studies on Viruses

The first X-ray studies on viruses were performed as early as in 1941 (8). The early fascination was to ascertain the existence of a precise molecular structure for these particles capable of multiplication in a suitable host cell. The helical parameters of TMV were obtained from X-ray diffraction studies on its oriented gels (9). Development of techniques in electron microscopy revealed the subunit packing arrangements in several spherical viruses (10). The high resolution structure of an isometric virus particle (11) was obtained only after major advances were made in techniques of recording X-ray diffraction photographs and computational crystallography. Virus crystals lose their ability to diffract on prolonged exposure to radiation and hence it was necessary to design high flux X-ray sources to record diffraction data within short exposure times. Due to large unit cell sizes, photographically recorded reflection spots did not have the spatial resolution that permit reliable intensity estimates. Introduction of Franck's optics (12) for further focussing the X-ray beam from a fine focus rotating anode X-ray generator permitted the resolution of diffraction maxima from lattices up to 1000 Å periodicity. Conventional X-ray diffractometers or photographic methods exclude from measurement a large number of reflections potentially active at any crystal setting. Use of screenless oscillation photography alleviated this problem and allowed rapid accumulation of diffraction data. However, the resulting photographs

were forbiddingly complex for manual interpretation. Hence, rotating drum scanners to convert the photographs to a digital image suitable for automated processing were developed. Computer programs for the analyses of these images were written (13). Isomorphous replacement has been the method of choice for the structure determination of proteins. However, the same method, when applied to the structure of viruses, yielded comparatively poorer electron density maps needing further improvement for definitive interpretation. The poorer quality of the electron density map is a result of the weaker diffraction from virus crystals based on large unit cells. The method of molecular replacement (14) provided the means of further phase refinement. This method improves the phases by imposing equivalence of electron density in chemically equivalent, but crystallographically independent, regions of the crystal. Structures of several spherical viruses have been successfully determined in the past decade (11, 15–24). Recently, use of synchrotron radiation and supercomputers have reduced the time required for virus structure determination by at least an order of magnitude. These structures provide detailed information on protein–protein and protein–nucleic acid interactions, nature and extent of quasi-equivalent bonding, mechanisms of assembly and disassembly, evolutionary relationships and other aspects of virus life cycle.

The Structure of Rod-shaped Viruses

There are seven biologically distinct, ssRNA containing plant viruses with an elongated, rod-like morphology. TMV is the type member of tobamo viruses and is found naturally throughout the world. It has the appearance of a rigid rod with a stain penetrated central channel in electron micrographs. In contrast, potato virus X (PVX), potato virus Y (PVY) and carlaviruses consist of flexuous rod-shaped particles. Rod-shaped viruses do not form single crystals. Hence, structural studies on these viruses have to contend with well oriented fibres.

TMV has long been a favourite model for structure and assembly studies (25, 26). TMV particles are about 3000 Å long and 180 Å in diameter and have a rigid cylinder-like morphology. The protein coat consists of 2,130 identical subunits of molecular weight (M_r) 17,420 daltons. The protein subunits are arranged as a helix of pitch 23 Å with $16 \frac{1}{3}$ subunits per helical turn. An ssRNA molecule of 2×10^6 daltons is encapsulated in the protein shell. The course of the nucleic acid follows the helical arrangement of the protein subunits. The particle has a central hole of 40 Å as revealed by the penetration of negative stain in electron micrographs.

TMV protein subunits aggregate to form highly polymorphic oligomeric structures depending on the ionic strength and the pH of the solvent (25, 26). One of the aggregates formed at neutral or higher pH is the two-layered protein disk. Each layer of the disk consists of 17 protein subunits organized with rotational 17-fold symmetry. This aggregate appears to be the most important intermediate in the assembly of the virus. The X-ray structure of the crystals of the disk aggregate has been determined at 2.8 Å resolution (27). The final electron density map reveals that the oligomer obeys the 17-fold symmetry within experimental errors. The central part of each subunit consists of four closely parallel (or antiparallel) alpha helices at a distance of 45–65 Å from the centre. The further ends of helices are connected by short antiparallel beta-structure. The loops connecting the helical segment at the end proximal to the disk axis are partially disordered. In contrast, most of the protein subunit is well ordered in the 2.9 Å

structure of the intact virus particle obtained by fibre diffraction methods (26). The lateral inter-subunit contacts in the disk-like aggregate and the virus are similar (26, 27), except for the disorder of the loops on the inner side. The residues interacting at the inter subunit contact are generally similar, although, the subunits of the viral helix are tilted with respect to the orientation of the subunit in the disk. However, the axial inter-subunit interactions are completely different. The axial contacts are closer in the virus than in the disk.

Structure of Spherical Plant Viruses

The first high resolution structure of a spherical virus to be determined was that of tomato bushy stunt virus (TBSV) (11). Packing of protein subunits in TBSV capsid is based on a $T=3$ icosahedral lattice. Each of the 60 icosahedral asymmetric units consist of 3 quasi-equivalently related subunits designated A,B,C (figure 3). The 386-residue long polypeptide of TBSV coat protein folds into two distinct domains, the shell or S-domain and the protruding or P-domain. The S-domain (residues 104–125) consists of an 8-stranded antiparallel β -barrel with a topology which has been described as “swiss roll” or “jelly roll”. The S-domain has a wedge shape which allows efficient packing at the 5-fold and quasi 6-fold axes without leaving large voids. The wedge shape also results in a large surface area of inter-subunit contact minimizing the free energy of interaction between subunits and hence imparting stability to the assembled particle. The P-domain (residues 273–386) consists of 10 β -strands with a topology similar to that of the S-domain. It has been, therefore, suggested that the two

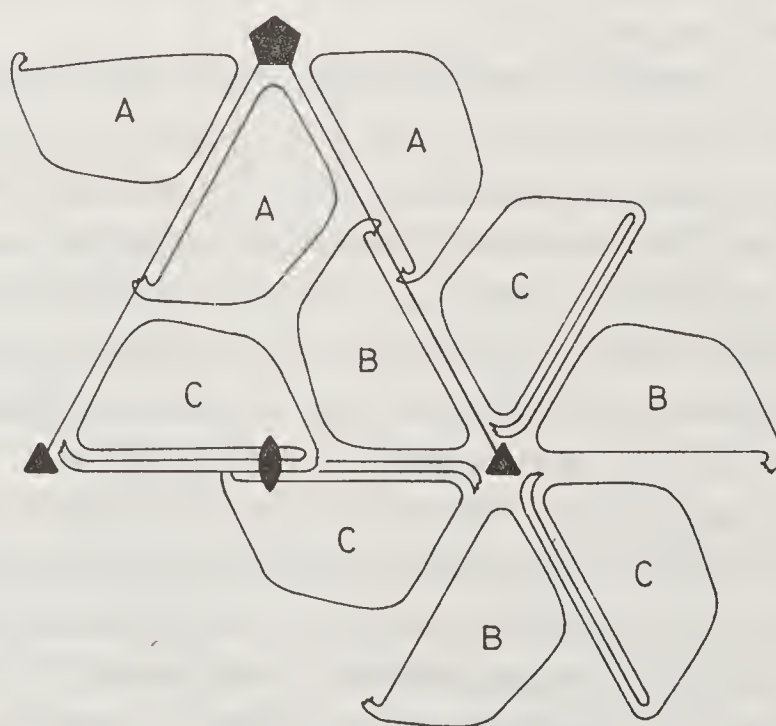


Figure 3. Arrangement of the quasi-equivalent A, B and C subunits in TBSV and SBMV. The amino termini in A and B subunits are disordered and are shown as short knobs. The C subunit termini are ordered and interact at the icosahedral three-fold (quasi six-fold) axis to form the so-called “ β -annulus” structure. The A, B and C subunits are chemically identical in TBSV and SBMV. In CPMV, two domains of a larger subunit and another smaller subunit occupy positions equivalent to those of the C, B and A subunits of SBMV. Reproduced with the kind permission of Prof. M. G. Rossmann.

domains have probably resulted from a gene duplication. Residues 67–103 at the amino terminus of the polypeptide are disordered in the A and B subunits but well ordered in the C subunits. The ordered arms of three C type subunits pack between B and C subunits at the quasi 6-fold axis giving rise to the so-called β -annulus structure. Observation of this structure led to the proposal of an assembly mechanism in which a large $T=1$ structure with C type subunits is initially formed. In this structure all contacts are exactly equivalent. Addition of further subunits to this preformed particle can only take place with subunits in an altered conformation, explaining the lack of exact equivalence of the subunits.

TBSV particles have two bound calcium ions per protein subunit. Upon removal, the virus expands reversibly when the pH is raised above neutrality. The structure of the expanded virus has been determined to 8.0 Å resolution (28). The structures of the intact and expanded particles have also been examined by neutron scattering (29). In the swollen virus, the tertiary structure of the coat protein S- and P-domains are similar to those of the intact virus. However, the protein subunits move radially out resulting in the formation of channels about 80 Å long and 20 Å in radius at the quasi-three fold axes.

The structure of turnip crinkle virus (TCV) has been determined at 3.2 Å resolution (30). TCV is a member of the tombus virus group and has 23% amino acid sequence identity with TBSV coat protein. Not surprisingly, the structure of TCV is remarkably similar to that of TBSV.

The structure of southern bean mosaic virus (SBMV) was determined soon after that of TBSV (15). The protein subunits of SBMV (28 kd) are much smaller than those of TBSV (42 kd). Hence, it came as a great surprise that the two structures were exceedingly similar. SBMV lacks the P-domain. However, it has an S-domain (figure 4) and β -annulus structure organized in a manner very similar to that of TBSV. The unrefined coordinates of 179 alpha carbon atoms in the S-domain of SBMV might be superposed with structurally equivalent residues of TBSV with an r.m.s. difference of 2.2 Å (31). The structural components are superposed without any rotations or translations when the symmetry axes of the two icosahedral particles are aligned.

Upon trypsinization, the coat protein subunits of SBMV lose the amino terminal 61 residues resulting in a 22 kd polypeptide. These subunits assemble into $T=1$ icosahedral empty shells. The inter-subunit contacts at the 5-, 3- and 2-folds are similar to the contacts at the 5-fold, quasi 3-fold and quasi 2-fold contacts of the $T=3$ SBMV particle (32).

Satellite tobacco necrosis virus (STNV) is the smallest of the known plant viruses. It multiplies only when co-infected in the presence of a helper virus, tobacco necrosis virus. The particle architecture is based on a $T=1$ lattice with 60 copies of 20 kd protein subunits organized with icosahedral symmetry. The structure of STNV has been determined at 2.5 Å resolution (16). Surprisingly, the tertiary structure of the polypeptide was found to resemble the S-domain of TBSV and SBMV. However, the quaternary structure of the pentamers of STNV and SBMV are significantly different. The five-fold axes of STNV and SBMV make an angle of 19° when a pentameric subunit of STNV is superposed on a similar subunit from SBMV.

The three-dimensional structures of two comoviruses, cowpea mosaic virus (CPMV) and bean pod mottle virus (BPMV) have recently been determined (18, 19). These viral capsids have a $T=1$ structure. The icosahedral asymmetric unit consists of two polypeptides of M_r 42 and 24 kd, respectively. The tertiary structures of these

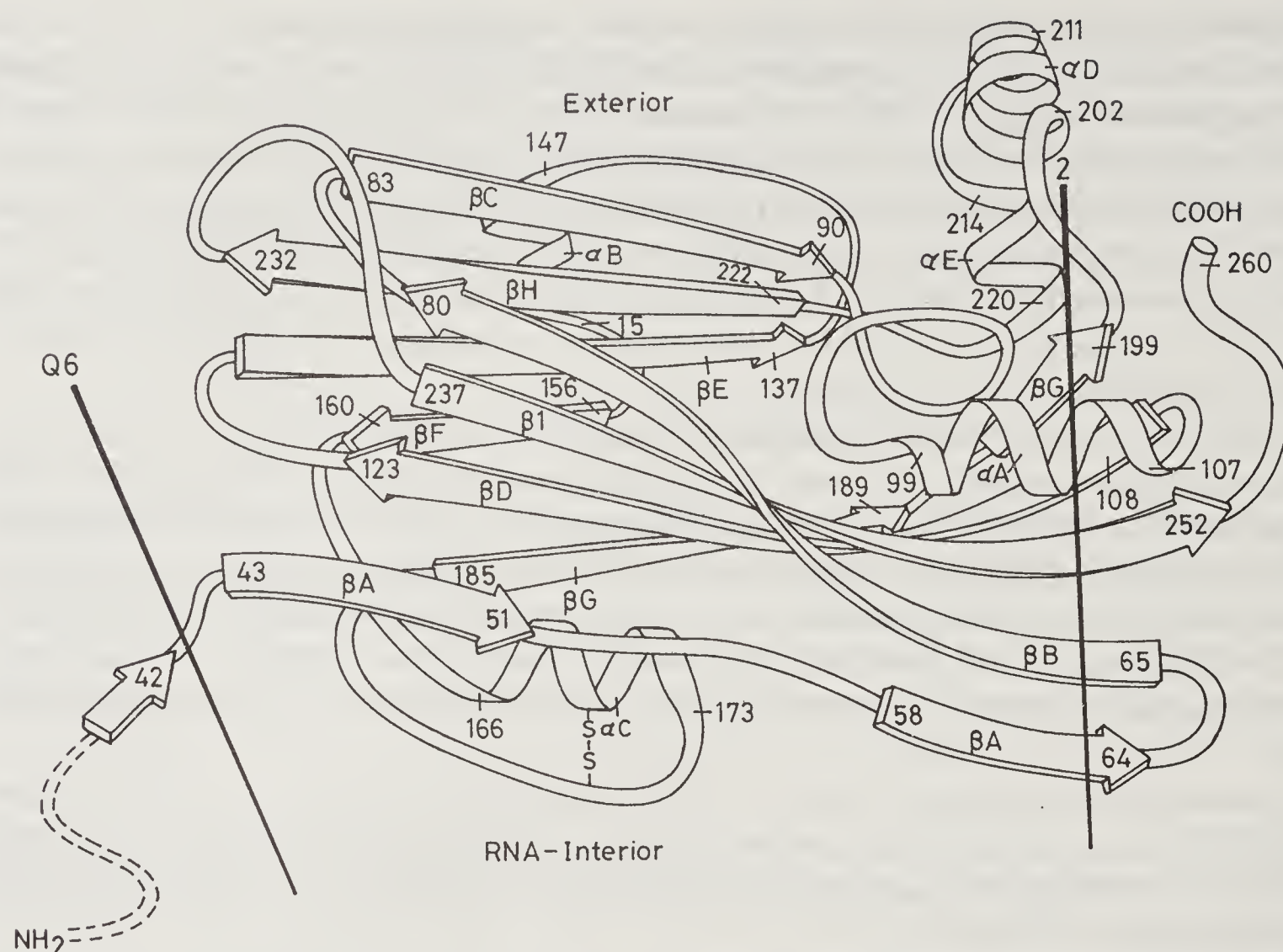


Figure 4. Diagrammatic representation of the S-domain of the SBMV coat protein. The tertiary structure of coat proteins of all isometric viruses except that of bacteriophage MS2 show a striking resemblance to this structure. Reproduced with the kind permission of Prof. M. G. Rossmann.

polypeptides consist of 2 and 1 β -barrel motifs, respectively, similar in topology to the S-domain of SBMV. It is likely that the 42 kd protein is a product of gene duplication from an ancestral gene coding for a β -barrel protein. These three domains are organized in a pseudo $T = 3$ arrangement, reminiscent of SBMV structure. One of the most intriguing observations of these studies is the partially ordered nucleic acid structure suggesting at least partial icosahedral symmetry and an ordered orientation of the nucleic acid. Nearly 20% of the BPMV RNA is visible in the electron density maps. The visible portion of RNA is single-stranded with easily recognizable polarity and stereochemistry. These results suggest that specific regions of the coat protein interact non-specifically with the asymmetric nucleic acid imparting icosahedral symmetry and rigid structure to RNA. In the high resolution electron density maps of other viruses, no discernable feature is visible for the RNA region. The basic amino terminal segments of TBSV, SBMV and STNV inferred to interact with nucleic acid are also disordered. However, a patch of positively charged residues are found clustered in the interior strands of the β -barrel facing the nucleic acid in SBMV. Probably, this patch of residues is responsible for specific recognition of the cognate nucleic acid by the protein coat.

Belladonna mottle virus (BDMV) belongs to the tymovirus group of spherical plant viruses (33). BDMV is characterized by particles of 300 Å diameter and multiple components containing different lengths of RNA in natural purified preparations. However, in the case of the related turnip yellow mosaic virus (TYMV), the full length

genome is sufficient to cause infection. The amino acid sequence of the 20 kd coat protein of BDMV-P strain has recently been determined (34). Comparison of the amino acid sequence of BDMV with TMV (35) and egg plant mosaic virus (36, EMV), the two other tymoviruses of known sequence, shows that BDMV is closely related to EMV while TYMV is evolutionarily equidistant from BDMV and EMV. No recognizable amino acid similarity is found between tymoviruses and other plant viral coat protein sequences. A conspicuous feature of these sequences is the absence of a basic amino terminal segment which is a common feature of several other plant viruses. These basic arms are responsible for partial neutralization of the negative charges on RNA phosphates and are probably involved in the initiation of the assembly of virus particles (37, 38). The absence of the basic amino terminal segment in tymoviruses is correlated with the presence of polyamines bound to the nucleic acid. These polyamines can easily be replaced by other monovalent cations in BDMV. Replacement of polyamines by monovalent cations leads to instability of the BDMV particles at alkaline pH and is associated with the release of RNA (39, 40). From these observations, it may be concluded that the polyamines not only partially neutralize the RNA phosphates but also impart specific stability properties. Low angle neutron scattering studies indicate that the protein subunits of TYMV are densely packed with little, if any, penetration of the nucleic acid into the protein coat (41). This is in contrast to cucumber mosaic virus, where the stability properties are dominated by strong protein–nucleic acid interactions. Hence, TYMV is regarded as the prototype example of viruses stabilized by strong hydrophobic association of protein subunits.

BDMV has been crystallized in rhombohedral space group $R3$ with cell edges of 295.38 Å and interaxial angles of 59.86°. Using 5.5 Å resolution X-ray diffraction data, the icosahedral symmetry of the virus particles and their orientation in the unit cell have been deduced (42). Cross-rotation function studies with SBMV and CPMV revealed significant structural similarities between tymo viruses and other $T=3$ plant RNA viruses of known three-dimensional structures, despite the lack of amino acid sequence similarities (43). Using SBMV structure as a starting model, the structure of BDMV was determined at 5.5 Å resolution by molecular replacement phase refinement. The resulting electron density map reveals the presence of large holes at the 3- and 5-fold axes of the icosahedral particles probably responsible for exchange of polyamines.

AMV particles are probably based on helically symmetric protein closed on either side by icosahedral caps. The length of the particles is determined by the size of RNA molecule encapsidated. Structure of the reconstituted $T=1$ AMV has been determined to 4.5 Å resolution (44). Though the course of the polypeptide chain could not be traced, there are indications that the structure of the polypeptide is similar to the β -barrel structure found in other viruses.

Structure of Animal and Insect Viruses

The first animal virus structure to be determined is that of human rhinovirus 14 (20). This was soon followed by the structure determination of polio virus (22). More recently, the structures of mengo virus (23) and foot and mouth disease virus (FMDV, 24) have been determined. These viruses belong to the picorna (small RNA) class of

animal viruses causing disease in a diverse class of animals including humans. The polypeptide folds of these viruses are similar, although there are some important differences. In these viruses, the 60 identical icosahedral asymmetric units consist of 4 polypeptide chains designated as VP1, VP2, VP3, VP4. These proteins do not display any significant sequence homology. Each of the major capsid proteins, VP1, VP2, VP3 consist of an 8-stranded antiparallel β -barrel similar to the S-domain of spherical plant viruses. The three protein domains are related by a quasi 3-fold axis resulting in a quaternary arrangement similar to the A, B, C subunits of SBMV. Hence, the surface lattice of protein domains might be described as pseudo $T=3$. It appears that the chemical equivalence of subunits is not required for the formation of lattice types predicted on the basis of the quasi-equivalence hypothesis.

Critical examination of the picorna viral structures has suggested a strategy of evasion of immune surveillance by these viruses (45–46). The putative receptor binding site in HRV14 and mengo virus might be described as a canyon or pit, respectively. The dimensions of these binding sites allow recognition of smaller cellular receptors but sterically hinder larger antibody molecules. Mutation of residues at the rim of the canyon or pit is believed to allow evolution of new serotypes without changes in receptor specificity. The structure determination of HRV14 complexed with antiviral compounds that inhibit uncoating of the virus has suggested a molecular mechanism of drug action and holds promise for rational drug design to combat widespread diseases caused by picornaviruses.

Black beetle virus (BBV) is the only insect virus whose three-dimensional structure has been determined (47). The protein subunits are organized on a $T=3$ surface lattice. The tertiary structure and packing of the subunits is reminiscent of plant and animal viruses. The largest differences between these structures is confined to the loops connecting the strands of the β -barrel.

The structure of bacteriophage MS2 has recently been determined at 3.2 Å resolution (48). The structure is dissimilar to other spherical viruses. This is the first ssRNA containing icosahedral virus built of protein subunits without the 8-stranded β -barrel domain. Each icosahedral asymmetric unit of MS2 coat consists of three chemically identical protein subunits designated as A, B and C. Each monomer has the shape of a triangle with a 60 Å base, 20 Å height and 20 Å thickness. The RNA facing side consists of 5 β -strands with a meander (antiparallel strands connected by a short loop) topology. The exposed part of the subunit consists of two or more strands and two helices at the carboxy terminus. The structure resembles those of bovine platelet factor 4 and a domain of class I histocompatibility antigen HLA-A2.

Polyoma virus particles are icosahedral with a mean diameter of 490 Å. The genetic material is a dsDNA (3.3×10^6 d) encapsidated within a shell composed primarily of the major coat protein VP1 (42 kd) and the minor proteins VP2 (35 kd) and VP3 (23 kd). Klug showed the presence of 72 morphological units whose positions confirmed to a $T=7$ surface lattice (49). The structure of polyoma virus has been determined to 22.5 Å resolution (50–52). The substructure of the particle revealed the first major violation of the theory of quasi-equivalence. According to the theory, it was anticipated that the virion would have 60 hexameric and 12 pentameric capsomeres on the surface of the particle. However, the low resolution structure indicates pentameric substructure for all the 72 morphological units. Thus the capsid has only 360 subunits, instead of 420 as predicted by Klug (49). These violations of quasi-equivalence hypothesis have prompted suggestions for other mechanisms of error-free assembly.

Virus Assembly

Much of the information on virus assembly has been derived from *in vitro* studies. Under suitable conditions of ionic strength, temperature, and pH, capsid proteins of some viruses assemble with their cognate nucleic acids to form virus-like particles. The assembly process is complete if the reassembled particles are indistinguishable from native virus particles in terms of their infectivity, stability *in vitro*, and hydrodynamic properties. Similar self-assembly processes might be central to the organization of other assemblies of macromolecules such as ribosomes or multi-enzyme complexes. Hence, the self-assembly mechanisms of several viruses have been extensively investigated. The assembly of TMV proceeds via discrete intermediates (25). The 17-fold symmetric, 2-layered protein disk is a well characterised intermediate. A base paired loop structure in RNA approximately 1000 nucleotides downstream from the 3' terminus inserts into the central cavity of the disk and initiates assembly. Both the 3' and the 5' ends of RNA protrude from the same end of this initiation complex. Binding of RNA brings about a change in the inter-protein subunit contacts from a stacked disk-like structure to a helix-like structure. Further assembly takes place by fast addition of disks to the free end while RNA is pulled up through the central hole. As 3 nucleotides interact with each protein subunit, approximately 100 nucleotides are protected from nuclease attack by the addition of each ring. Only the initiation step appears to have specificity as the assembly can proceed normally with other RNA's to which the loop-like structure has been engineered (53). In comparison, the mechanisms of assembly of spherical plant viruses is far less clear. Assembly of protein and nucleic acid components into complete virions has been so far possible only with those viruses which are predominantly stabilized by protein-nucleic acid interactions.

The disassembly process in bromoviruses has been studied in considerable detail (54). The stability of the virus particles is predominantly due to protein-nucleic acid interactions. The virus particles sediment with an S value of 88 at pH 5.0. On raising the pH to above 6.5, a large number of protons are released and the sedimentation coefficient drops to 78S indicating a swelling of the virus particles. The increase in the hydrodynamic radius of the virion is also confirmed by low angle neutron scattering measurements (55). The swelling process can be reversed by lowering the pH. However, compaction of the particles and the concomitant increase in sedimentation coefficient occurs at a lower pH when compared to swelling, following a reproducible hysteresis loop. The encapsidated nucleic acid in the swollen virus is susceptible to ribonuclease. On removing divalent cations by treatment with EDTA and increasing the ionic strength to 2.0, the virus particles dissociate and the nucleic acid precipitates. Based on these observations, and studies on reversible swelling in the presence of varying concentrations of calcium, it has been suggested that the stability of the virus is dependent on hydrogen bonding between protonated carboxyl groups. The microenvironment, such as proximity of other carboxyl groups or RNA phosphate groups, lead to abnormally high pH values at which they are deprotonated. Roughly 2 abnormally titrating groups in the pH range, where brome mosaic virus (BMV) and cowpea chlorotic mottle virus (CCMV) swell, have been observed by titration experiments (56). On raising the pH, the repulsion between the deprotonated carboxyls leads to swelling. Addition of divalent cations neutralizes the charges on the carboxyls and reduces the swelling resulting in the increase of S value to 85. Removal of divalent cations under

these conditions severs the weak residual forces holding the capsid and disassembly ensues.

Several isometric viruses contain an amino terminal segment of 50–60 residues rich in lysines and arginines. The removal of 61 amino terminal residues from the coat protein of SBMV by treatment with trypsin yields 22 kd polypeptides which assemble only into $T = 1$ particles devoid of RNA (37), implicating the role of the amino terminal arm in assembly of $T = 3$ particles. The involvement of the basic amino terminal arm is also suggested by the structure of TBSV and SBMV where the arm from the 3 C-type subunits interdigitate to form the so-called β -annulus structure (11, 15). The assembly of TBSV was visualized to proceed through an intermediate, cage-like structure composed of C-subunits. All the 60 subunits of this loose capsid are equivalent and the structure is stabilized by the formation of the β -annulus. The intervening gaps between the C-type subunits allow additional subunits only in the A/B conformation. In this mechanism, the $T = 3$ structure is a result of the assembly pathway and no “molecular switch” is needed. However, this mechanism is not compatible with the *in vitro* assembly studies on SBMV. The purified RNA of SBMV is heterogeneous and can be separated into low and high molecular weight fractions on sucrose density gradients. Assembly of protein with high M_r RNA at pH 7 and pH 9 leads to $T = 3$ particles. Low M_r RNA assembles with protein into $T = 1$ particles at pH 5 and pH 7 and $T = 3$ particles at pH 9 (37). Rossmann (57) has formulated an hypothesis on the sequence of events during the self assembly of the $T = 3$ native virus particles and $T = 1$ empty capsids of SBMV. The coat protein is believed to exist in two states, “relaxed” AB type dimer and “tensed” CC type dimer. The amino terminal arms of relaxed dimers are disordered while those of tensed dimers are ordered. Transition from the “relaxed” to “tensed” state might be initiated by binding of ligands such as calcium. Initially five “relaxed” AB dimers might assemble to give rise to a 10-mer nucleation centre. Addition of further AB dimers can occur in two ways depending upon the state of protonation or Ca^{2+} ions binding to the carboxyl clusters in subunit contact regions. In low charge states, the “relaxed” AB dimers nucleate a $T = 1$ capsid and in high charge states, “tensed” CC dimers bind to the nucleation centre, resulting in a hydrophobic cleft between B and C subunits. This would lead to the formation of the β -annulus, and a $T = 3$ capsid. The pathway of assembly is thus controlled by molecular switches whose state is altered by ligand binding.

Picornavirus assembly proceeds through discrete intermediates (58). The capsid proteins VP0, VP3 and VP1 initially assemble into a 6S protomer. Five such protomers cluster with 5-fold symmetry to form a 14S intermediate which then assembles into the procapsid. Either during the assembly or soon after the formation of the capsid, the RNA is encapsidated resulting in the simultaneous cleavage of VP0 into VP2 and VP4. The dissociation of mature virion, therefore, does not provide the molecules responsible for *in vivo* assembly.

Polyoma virus capsids are built of only pentamers of the major capsid protein VP1 (51). The $T = 7$ icosahedral surface lattice is occupied by 12 pentavalent and 60 hexavalent pentamers instead of 12 pentameric and 60 hexameric capsomeres as anticipated by the theory of quasi-equivalence. These observations have prompted a careful re-examination of the assembly of polyoma virus (59–60). The VP1 gene has been cloned and the protein has been purified from the expressing clone of *E. coli*. The assembly experiments with the protein obtained from the recombinant gene reflect the intrinsic propensities of the VP1 protein as any “switching” by other molecules that

might be interfering with the assembly process during replication *in vivo* is absent. The observation that this protein also leads to only pentameric capsomeres suggests that the non-equivalent bonding pattern observed in the polyoma virus structure is due to the inherent properties of the VP1 molecule. It has been found that the state of aggregation of the VP1 protein depends not only on the final solvent conditions, but also on the pathway of solvent changes.

Interference with the assembly/disassembly of virus particles provides a method for combating viral diseases. For example, transgenic plants expressing TMV or CMV coat proteins are found to be resistant to infection by the respective viruses as the presence of excess coat protein inhibits virus disassembly (61).

Virus Evolution

Plant viruses have been classified into a large number of groups based on several of their characteristics, like particle morphology, host range, vector specificity, segmented or non-segmented genomes etc. (1). Serological cross reactivity has been used to establish the evolutionary relationship of some of these classes exhibiting comparable biological properties such as CCMV and BMV. Recently, the genomic sequence of several viruses have been determined. These sequences provide a wealth of information on the evolutionary history of these viruses. The comparison of the RNA3 sequence of CMV and BMV illustrates several of the important factors of viral evolution (62). RNA3 in these tripartite viruses code for 2 proteins, the coat protein and another non-structural protein. The non-structural proteins have an overall amino acid sequence similarity of 33%. Another 50% of the aligned residues are strictly conservative substitutions. These observations clearly establish the evolutionary relationship of these two viruses. It is interesting to note that Kaper (63) had classified these viruses into a single group solely based on the similarity of the chemical and stability properties of these viruses. Both the viruses degrade readily in the presence of high concentrations of neutral chloride salts and SDS. Also the encapsidated RNA is sensitive to ribonuclease. In contrast to the striking similarity of the 3a proteins of these viruses, the capsid proteins do not have any significant similarities. High and low degree of similarities has also been observed in the non-structural and structural proteins, respectively, of other viruses (64). These studies suggest that the non-structural proteins are under stringent evolutionary constraints in comparison with capsid proteins. Further, the relatively faster rate of change in the coat protein sequence suggests that the mechanisms by which virions assemble can undergo rapid evolutionary adaptations in comparison with strategies of replication. Also, the 3'-terminal non-coding sequences of CMV and BMV can be folded into a t-RNA like structure. This region is also highly conserved suggesting its importance in the life cycle of these viruses (62). Amino acid homology has also been observed in the virally coded polymerase sequences of animal picornaviruses and plant CPMV (65). Based on similar observations on many classes of viruses, Goldbach and Wellink have suggested that the ssRNA viruses might be grouped into two major superfamilies. The picornavirus superfamily includes plant como-, nepo-, potyviruses, animal picornaviruses etc. These viruses have a protein covalently attached at the 5'-end to the viral nucleic acid. The genomic RNA, which has a poly A tail at the 3'-end, is initially expressed as a polyprotein and subsequently is proteolytically processed to yield functional proteins.

In contrast, the sindbis virus-like family includes the plant ilar, AMV, bromo-, cucumo-, tobamo-, tobra-, hordei-, carmo-, and tombusviruses. Although, these viruses display wide variation in genomic organization and mode of expression, the genomic RNA has a 5'-cap structure in all of them. The functional proteins are translated from sub-genomic messengers. Three non-structural proteins display significant homology. A low degree of sequence similarity has also been observed between the two superfamilies of viruses. The three dimensional structures of the isometric plant, animal and insect viruses display a remarkable similarity with the exception of the recently determined structure of the bacteriophage MS2. The tertiary structure of the coat proteins of all these viruses is similar and consists of the wedge shaped 8 stranded antiparallel β -barrel. Hence, it is likely that these viruses have evolved from a common ancestor. However, a divergent evolutionary mechanism alone is not sufficient to account for all the observed data. For example, TMV protein is rich in alpha helix in contrast to the β -structure of icosahedral viruses. Also the unique gene orders observed with some of these viruses is not compatible with divergent evolution as the sole mechanism of viral evolution (64). Independent incorporation of genes from host cell into viral sequences followed by gene rearrangements and other as yet unspecified mode of inter-viral RNA recombination probably account for the observed similarities and differences.

Acknowledgements

We thank Drs. Sanjeev K. Munshi and Dinakar M. Salunke for comments on the manuscript. Thanks are also due to the Department of Science and Technology, for financial support to MRNM and HSS.

References

1. Matthews, R. E. F. (1981) *Plant Virology*, 2nd Edn (Academic Press, New York).
2. Luria, S. E., Darnell, J. E. Jr., Baltimore, D. & Campbell, A. (1978) *General virology*, 3rd Edn (John Wiley & Sons, New York).
3. Crick, F. H. C. & Watson, J. D. (1956) *Nature (London)* **177**, 473–475.
4. Pauling, L. & Corey, R. B. (1951) *Proc. Natl. Acad. Sci., U.S.A.* **37**, 729–740.
5. Klug, A. (1979) *Harvey, Lect.*, **74**, 141–172.
6. Casper, D. L. D. & Holmes, K. C. (1969) *J. Mol. Biol.* **46**, 99–133.
7. Caspar, D. L. D. & Klug, A. (1962) *Cold Spring Harbor Symp. Quantum Biol.* **27**, 1–24.
8. Bernal, J. D. & Fankuchen, I. (1941) *J. Gen. Physiol.* **25**, 111–165.
9. Caspar, D. L. D. (1963) *Adv. Protein Chem.* **18**, 37–121.
10. Crowther, R. A. & Klug, A. (1975) *Annu. Rev. Biochem.* **44**, 161–162.
11. Harrison, S. C., Olson, A., Schull, C. E., Winkler, F. K. & Bricogne, G. (1978) *Nature (London)* **276**, 368–373.
12. Harrison, S. C. (1968) *J. Appl. Crystallogr.* **1**, 84–90.
13. Arndt, U. W. & Wonacot, A. J. (1977) *The Rotation Method in Crystallography* (North Holland, Amsterdam).
14. Argos, P. & Rossmann, M. G. (1980) *Molecular Replacement Method in Theory and Practice of Direct Methods in Crystallography*, eds. Ladd, M. F. C. & Palmer, R. A. (Plenum Press, New York).
15. Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D. & Tsukihara, T. (1980) *Nature (London)* **286**, 33–39.
16. Liljas, L., Unge, T., Jones, T. A., Fridborg, K., Lovgren, S., Skoglund, V. & Strandberg, B. (1982) *J. Mol. Biol.* **159**, 93–108.

17. Hogle, J. M., Maeda, A., Harrison, S. C. (1986) *J. Mol. Biol.* **191**, 625–638.
18. Stauffacher, C. V., Usha, R., Harrington, M., Schmidt, T., Hosur, M. V. & Johnson, J. E. (1987) in *Crystallography in Molecular Biology*, eds. Moras, D., Drenth, J., Strandberg, B. S. & Wilson, K. (Plenum Press, New York/London) pp. 293–308.
19. Chen, Z., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonosoff, G. & Johnson, J. E. (1989) *Science* **245**, 154–159.
20. Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H. J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., Rueckert, R. R., Sherry, B. & Vriend, G. (1985) *Nature (London)* **317**, 145–153.
21. Roberts, M. M., White, J. L., Grutter, M. G. & Burnett, R. M. (1986) *Science* **232**, 1148–1151.
22. Hogle, J. M., Chow, M. & Filman, D. J. (1985) *Science* **229**, 1358–1365.
23. Luo, M., Vriend, G., Kamer, G., Minor, I., Arnold, E., Rossmann, M. G., Boege, U., Scraba, D. G., Duke, G. M. & Palmenberg, A. N. (1987) *Science* **235**, 182–191.
24. Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D. & Brown, F. (1989) *Nature (London)* **337**, 709–716.
25. Butler, P. J. G. (1984) *J. Gen. Virol.* **65**, 253–279.
26. Namba, K., Pattanayek, R. & Stubbs, G. (1989) *J. Mol. Biol.* **208**, 307–325.
27. Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978) *Nature (London)* **276**, 362–368.
28. Robinson, I. K. & Harrison, S. C. (1982) *Nature (London)* **297**, 563–568.
29. Kruse, J., Kruse, F. M., Witz, J., Chauvin, C., Jacrot, B. & Tardieu, A. (1982) *J. Mol. Biol.* **162**, 393–417.
30. Hogle, J. M., Maeda, A. & Harrison, S. C. (1986) *J. Mol. Biol.* **191**, 625–638.
31. Rossmann, M. G., Abad-Zapatero, C., Murthy, M. R. N., Liljas, L., Jones, T. A. & Strandberg, B. (1983) *J. Mol. Biol.* **165**, 711–736.
32. Erickson, J. W., Silva, A. M., Murthy, M. R. N., Fita, I. & Rossmann, M. G. (1985) *Science* **229**, 625–629.
33. Harrison, B. D., Finch, J. T., Gibbs, A. J., Hollings, M., Shepherd, R. J., Valenta, V. & Wetter, C. (1971) *Virology* **45**, 356–363.
34. Suryanarayana, S., Appaji Rao, N., Murthy, M. R. N. & Savithri, H. S. (1989) *J. Biol. Chem.* **264**, 6273–6279.
35. Peter, R., Stehelin, D., Reinbolt, J., Collot, D. & Duranton, H. (1972) *Virology* **49**, 615–617.
36. Dupin, A., Peter, R., Collot, D., Das, B. C., Peter, C., Bouillon, P. & Duranton, H. (1984) *C. R. Hebd. Seances Acad. Sci.*, **C298**, 219–221.
37. Erickson, J. W. & Rossmann, M. G. (1982) *Virology* **116**, 128–136.
38. Savithri, H. S. & Erickson, J. W. (1983) *Virology* **126**, 328–335.
39. Savithri, H. S., Munshi, S. K., Suryanarayana, S., Diwakar, S. & Murthy, M. R. N. (1987) *J. Gen. Virol.* **68**, 1533–1542.
40. Suryanarayana, S., Jacob, A. N. K. & Savithri, H. S. (1988) *Indian J. Biochem. Biophys.* **25**, 580–584.
41. Jacrot, B., Chauvin, C. & Witz, J. (1977) *Nature (London)* **266**, 417–421.
42. Munshi, S. K., Hiremath, C. N., Savithri, H. S. & Murthy, M. R. N. (1987) *Acta Crystallogr.* **B43**, 376–382.
43. Hiremath, C. N., Munshi, S. K. & Murthy, M. R. N. (1990) *Acta Crystallogr.* **B46**, 562–567.
44. Fukuyama, K., Abdel-Meguid, S. S., Johnson, J. E. & Rossmann, M. G. (1983) *J. Mol. Biol.* **167**, 873–894.
45. Rossmann, M. G. & Palmenberg, A. C. (1988) *Virology* **164**, 373–382.
46. Rossmann, M. G. & Rueckert, R. R. (1987) *Microbiol. Sci.* **4**, 206–214.
47. Hosur, M. V., Schmidt, T., Tucker, R. C., Johnson, J. E., Gallagher, T. M., Selling, B. M. & Rueckert, R. R. (1987) *Proteins* **2**, 167–176.
48. Valegard, K., Liljas, L., Fridborg, K. & Unge, T. (1990) *Nature (London)* **345**, 36–41.
49. Klug, A. (1965) *J. Mol. Biol.* **11**, 424–431.
50. Rayment, I., Baker, T. S., Caspar, D. L. D. & Murakami, W. T. (1982) *Nature (London)* **295**, 110–115.
51. Rayment, I. (1984) in *Biological Macromolecules and Assemblies, Virus Structure*, eds. Jurnak, F. A. & McPherson, A. (John Wiley and Sons, New York). Vol. 1, pp. 225–298.
52. Rayment, I., Baker, T. S. & Caspar, D. L. D. (1983) *Acta Crystallogr.* **B39**, 505–516.
53. Turner, D. R., McGuigan, C. J. & Butler, P. J. G. (1989) *J. Mol. Biol.* **209**, 407–422.

54. Lane, L. C. (1981) in *Handbook of Plant Virus Infections and Comparative Diagnosis*, ed. Kurstak, E. (Elsevier/North Holland Biomedical Press, Amsterdam).
55. Chauvin, C., Pfeiffer, P., Witz, J. & Jacrot, B. (1978) *Virology* **81**, 419–432.
56. Pfeiffer, P. & Durham, A. C. H. (1977) *Virology* **81**, 419–432.
57. Rossmann, M. G. (1984) *Virology* **134**, 1–11.
58. Ruectert, R. R. (1976) in *Comprehensive Virology*, eds. Fraenkel-Conrat, H. & Wagnes, R. R. (Plenum Press, New York) Vol. 6, pp. 131–213.
59. Salunke, D. M., Caspar, D. L. D. & Garcea, R. L. (1986) *Cell* **46**, 895–904.
60. Salunke, D. M., Caspar, D. L. D. & Garcea, R. L. (1989) *Biophys. J.* **56**, 887–900.
61. Powell Abel, P., Nelson, R. S., De, B., Hoffmann, N., Rogers, S. G., Fraley, R. T. & Beachy, R. N. (1986) *Science* **232**, 738–743.
62. Murthy, M. R. N. (1983) *J. Mol. Biol.* **168**, 469–475.
63. Kaper, J. M. (1975) in *The Chemical Basis of Virus Structure, Dissociation and Reassembly*, eds. Neuberger, A. & Tatum, E. L. (North Holland/American Elsevier, New York, Amsterdam) pp. 1–485.
64. Goldbach, R. & Wellink, J. (1988) *Intervirology* **29**, 260–267.
65. Franssen, H., Leunissen, J., Goldbach, R., Lomonossoff, G. & Zimmern, D. (1984) *EMBO J.* **3**, 885–861.

Metabolism and Maturation Studies of Intact Spermatozoa Using Magnetic Resonance Techniques

Sudha Srivastava*, Elvyra J. Fernandes⁺ Umashashi C. Hegde⁺,
Ratna S. Phadke* and Girjesh Govil*

*Tata Institute of Fundamental Research, Homi Bhabha Road,
Bombay 400 005, India

⁺Institute for Research in Reproduction, Bombay 400 012, India

Introduction

Unlike other body cells, spermatozoa are endowed with two highly specialized markers of activity, namely, the capacity to move fast (motility) and the ability to fertilize. These provide unusual opportunities for correlating chemical and metabolic measurements with the cellular activity (1). Moreover, these cells depend on sugars for their energy requirements, which they utilize anaerobically as well as aerobically at a rate which is closely related to the degree of sperm motility (2).

Magnetic resonance techniques (^{31}P , ^{13}C NMR and ESR spin labelling) are unique as they are non-destructive and non-invasive by nature and can be used for the study of intact cells under a variety of biological conditions (3). The techniques can provide detailed information about the cell's viability, metabolic activity, intracellular events and environment. Spermatozoa can be conveniently studied using these techniques as they neither grow nor multiply and can survive for a long time outside the body. Due to their motile nature, cell-settling does not pose serious problems inside spinning NMR tubes unlike other cell lines.

We have used these techniques on different systems such as epididymal cells from goat, rat and human ejaculates. Here, we present some of our results on metabolism and maturation to indicate the potentialities of these techniques for studying intact sperm cells.

Experimentation

(a) *Sample Preparation:* Epididymis obtained from goats and rats were segmentally dissected into their caput, corpus and cauda regions. Each segment was mechanically dissociated by gentle mincing and tweezing followed by repeated gentle washing with buffer. The motility and concentration of the cells were checked from time to time using a light microscope. Sperms exhibiting motility higher than 60% were used for experiments.

NMR experiments were carried out on a Bruker AM 500 FT-NMR spectrometer.

For ESR experiments a 5-mM stock solution of 2,2,6,6-tetramethylpiperidine-N-oxyl (TEMPO) was prepared in Ringer solution. Aliquots were added to the sperm sample as and when required. The final concentration of spin label was maintained at 1-mM to avoid spin-spin interaction and consequent broadening of ESR signals. The procedure typically followed is described below.

A timer was started at the initial contact of sperms with the spin-label. The sample was thoroughly mixed and then transferred to 50- μ l capillary tubes. ESR spectra were recorded at regular time intervals on an X-band ESR spectrometer having a Varian 12-inch electromagnet and a Varian V-45601 100 kHz field modulation and detection unit. Appropriate controls were used to understand the influence of experimental conditions. The observations were normalized to 10^9 cells to enable comparative assessments. Experimental error was less than 1% in all experiments. The reversibility of spin label reduction was verified by adding external oxidizing agents to the control samples.

Results and Discussions

^{31}P NMR: ^{31}P NMR is advantageous for the investigation of cells, because it has a relatively wider range of chemical shifts and there are several important phosphate-containing metabolites such as ATP, ADP, etc. which are present in mM concentration in the cells (3). Since the cell membranes are impermeable to phosphorus metabolites, the spectra provide information regarding intracellular metabolism alone.

The ^{31}P NMR spectrum of goat spermatozoa obtained from the cauda region is shown in figure 1. The peaks are assigned by addition of various phosphate compounds to the cell extract. Phosphocreatine (PCr) was used as the external reference. Signals assigned are: adenosine triphosphate (ATP) giving rise to three peaks corresponding to α - β - and γ -phosphates at -8 , -16.5 and -3.5 ppm, respectively, and adenosine diphosphate (ADP) giving rise to α - and β -phosphate peaks at -8 and -3.5 ppm, respectively, which overlap with the α - and γ -peaks of ATP in that region.

Sperms show a reasonably high concentration of AMP giving rise to a signal at 5.5 ppm indicating a low adenylate energy charge (4), and an accelerated rate of ATP generating reactions. The intense peak at 2 ppm is due to inorganic phosphate (Pi) inside the cell and can be used to measure the pH of the system (5). In this case the position is indicative of an acidic pH (~ 6.6 pH). The next strong peak at 3.5 ppm corresponds to sugar phosphate (SP) and glyceryl phosphocholine (GPC) which are overlapping. GPC signals are insensitive to the pH. Spectra corresponding to the cells obtained from caput and corpus regions are also in figure 1. In these spectra, the adenine nucleotide phosphate signals are extremely weak and can be observed only with longer accumulation. On the other hand, the Pi shows slight downfield shift indicating increase of pH from cauda to caput region cells. It is difficult to comment about the signal at 3.5 ppm since GPC and SP signals overlap in this region. However, the signal shifts downfield from cauda to caput region cells because the signal of one of the components (SP) is sensitive to the pH. Noteworthy is the presence of the signal arising from phosphocholine (PC) at 1 ppm in the caput region which is weak in the other spectra (6). Figure 2 shows the spectrum for spermatozoa from the cauda region maintained for two hours in the buffer. At $t = 0$, 20 mg of fructose was added and the spectra accumulated for 11 minutes at different intervals of time. The Pi

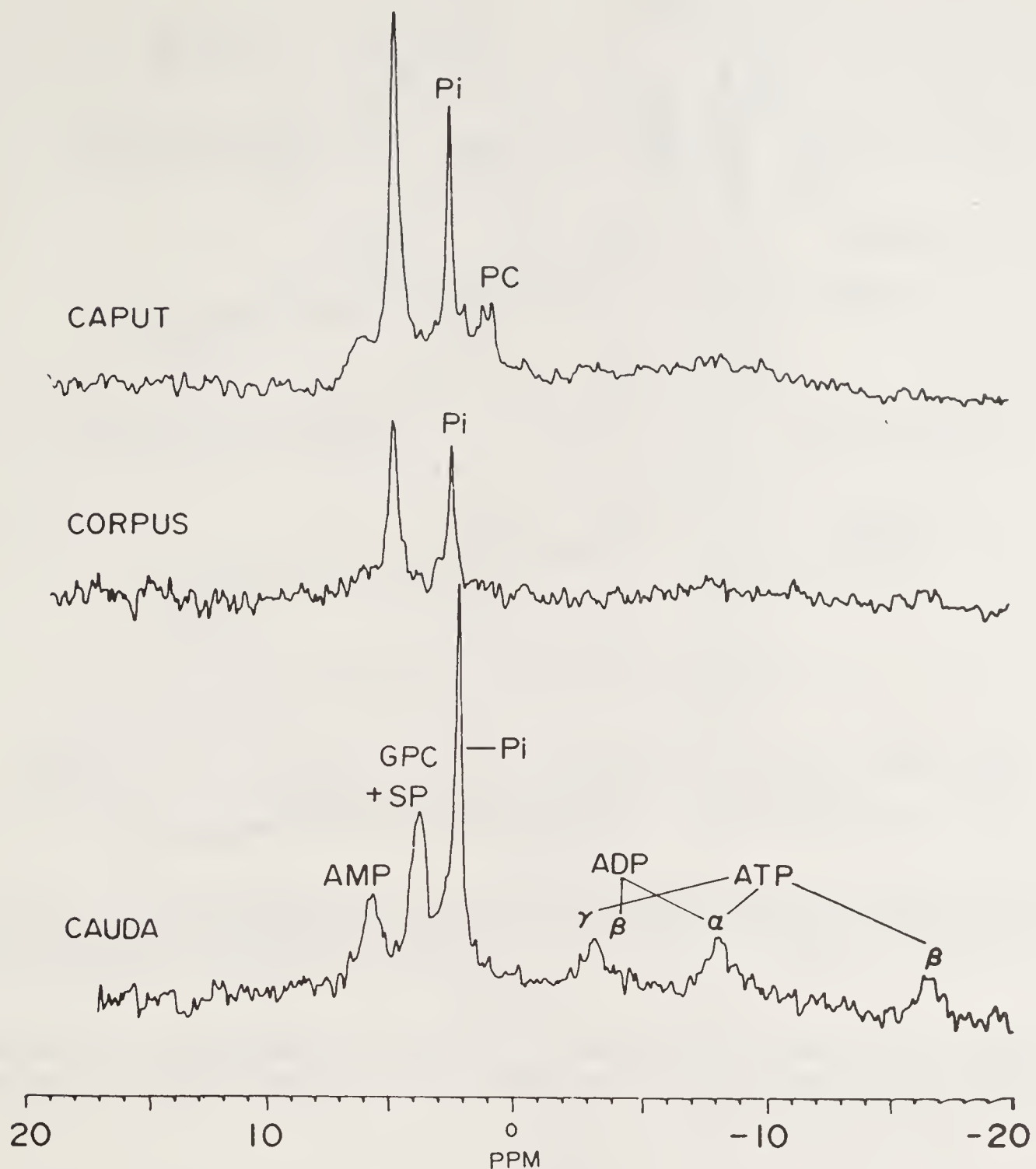


Figure 1. 202.15 MHz ^{31}P NMR spectra of goat spermatozoa obtained from caput, corpus and cauda regions of epididymis. Assignments have been indicated for all the signals.

signal undergoes upfield shift with time indicating an observable change in pH of the cells as a result of sugar metabolism. At $t = 0$, the sperms are quiescent having low ATP signals which build up gradually with time and persist even after 84 minutes. These observations provide evidence that the spermatozoa acquire the capability to withstand anaerobic fasting conditions and attain activation on addition of fructose.

Sperm capacitation in the female tract is associated with the shedding of the glycoprotein coat by spermatozoa (7). This in turn activates some of the enzymes essential for fertilization (8). To eliminate the female tract as the sole locus for capacitation, a variety of *in vitro* treatments have been devised to achieve the capacitation (9). We have used a medium which has close resemblance to the fluid of the female tract. Sperms collected from the cauda region were incubated in such a medium for different intervals of time and spectra were recorded (figure 3).

With increasing periods of capacitation one observes build up of ATP and sugar-phosphate signals at the expense of inorganic phosphate. It is also observed

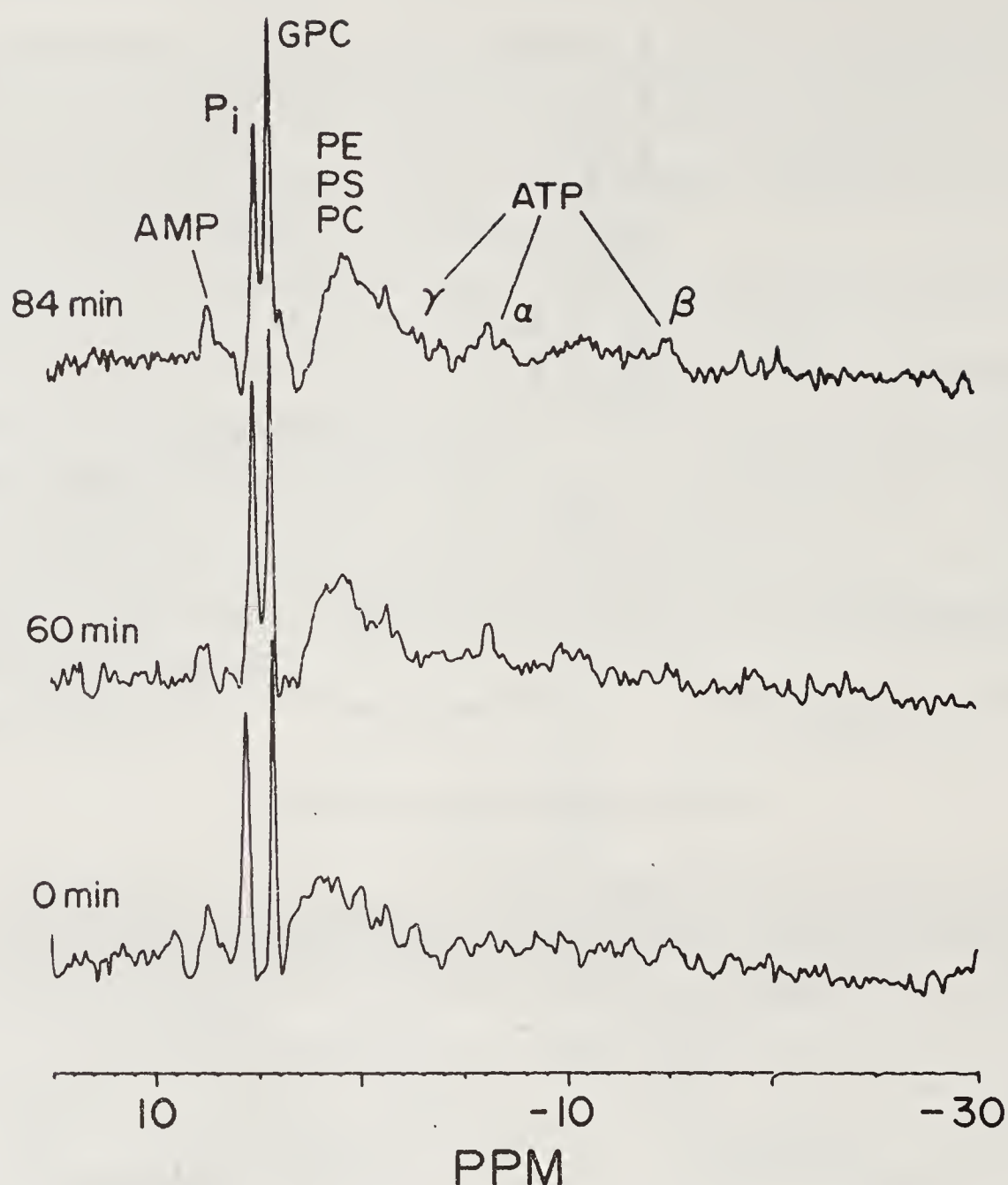


Figure 2. The sperms were starved for two hours by maintaining in the buffer without sugar. 20 mg of fructose was added at $t = 0$. The other spectra are shown for different time intervals.

that incubation beyond 6 hours is detrimental as the ATP levels do not remain the same. These results indicate that 3–6 h duration is optimum for the capacitation, in agreement with an earlier report (10).

^{13}C NMR: Using ^{31}P NMR, it is not possible to observe different sugar phosphates as they are too broad to allow separation and unambiguous assignment. Instead, ^{13}C NMR can be successfully used. It allows correlations between rates of glucose metabolism and levels of metabolites. Using ^{13}C labelled sugar as a substrate, it is possible to monitor most of the metabolites of glycolytic pathway in a single experiment. While accurate measurement of concentration is difficult and time-consuming, relative estimates can be made by comparing resonances arising from the same metabolites at different times.

Figure 4 shows the ^{13}C NMR spectrum of rat sperm cells obtained from the cauda region. 20 mM of glucose, labelled at C-1 position has been used as the substrate. The strong signals observed at 97 and 100 ppm are due to α - and β -isomers of glucose, respectively. The peak at 22.5 ppm is from lactate C-3 which is produced by the anaerobic oxidation of pyruvate. This signal shows an increase in signal intensity with time. The signal at 182 ppm corresponds to CO_2 which indicates that in addition to lactate, other products of pyruvate oxidation are acetic acid and carbon dioxide

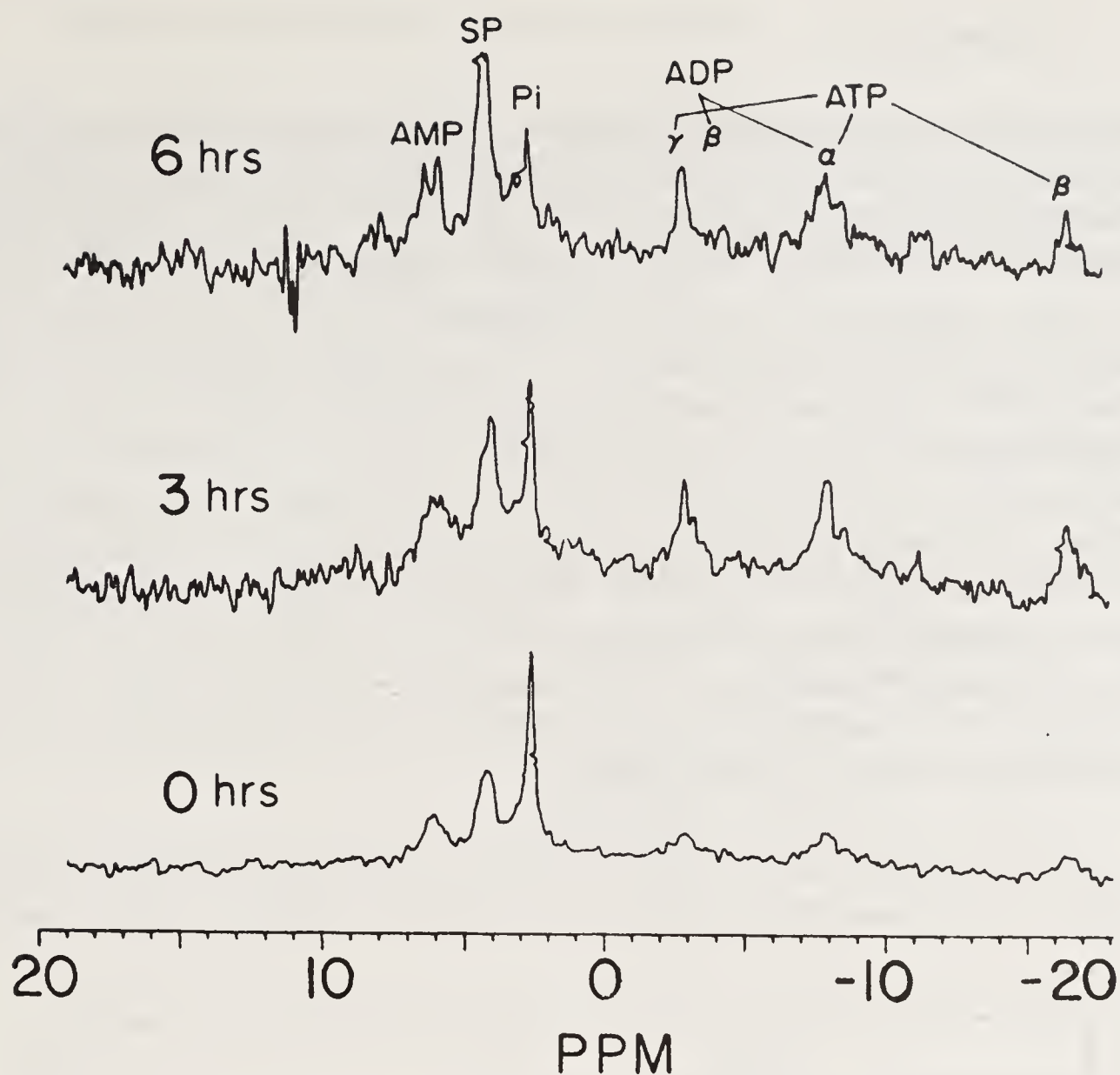


Figure 3. ^{31}P NMR spectra for spermatozoa obtained from cauda region of goat epididymis at different time intervals under *in vitro* capacitation. Experimental conditions are discussed in the text.

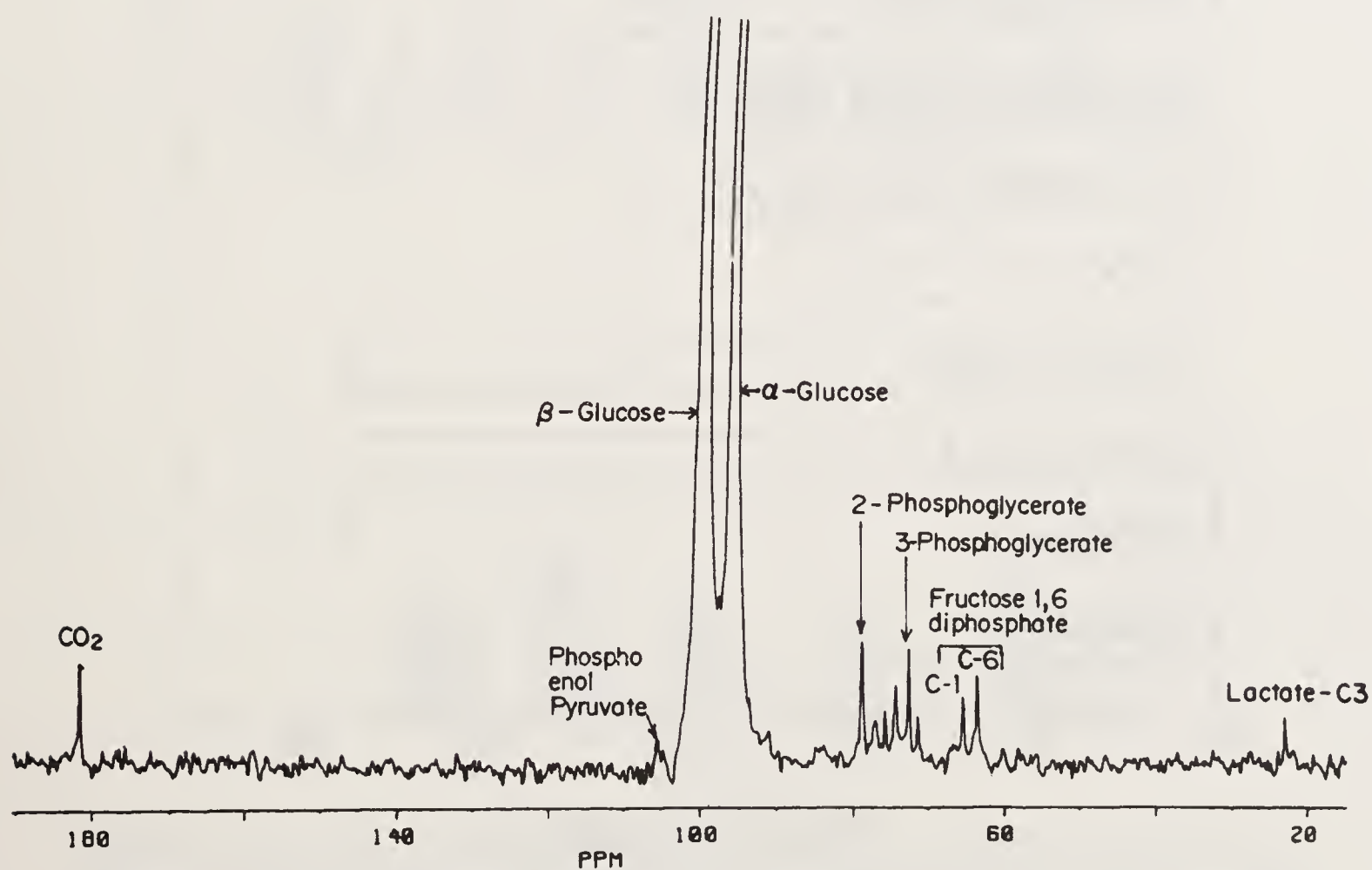


Figure 4. ^{13}C NMR spectrum of rat spermatozoa obtained from cauda region of epididymis incubated with 10 mg of 1- ^{13}C glucose. TSP has been used as external reference.

produced as a result of dismutation reaction. Another feature of the spectrum is a large number of resonances between 60 and 80 ppm. These usually arise from ^{13}C linked directly to hydroxyl groups. Some of these peaks could be assigned by adding the metabolites to the extract for locating the position of the corresponding ^{13}C resonance. Such assignments have been indicated in the figure. The appearance of these signals in the early stage of glucose ^{13}C -1 utilization precludes the presence of hexokinase, phospho-glucose isomerase, phospho-fructokinase and fructose diphosphate aldolase type of activities. Signals at 64 and 66 ppm corresponds to fructose 1,6 diphosphate, for both ^{13}C -6 and ^{13}C -1 atoms respectively. The presence of these signals shows scrambling of the label at the triose-phosphate isomerase stage. With this information, we plan to study cells from different regions of the epididymis to pin-point the metabolic differences, which in turn, can be used as markers for sperm maturation. Moreover, working at low glucose concentration, and with careful experimentation, it is possible to use the rate of glucose disappearance and lactic acid production for assay of semen quality.

ESR: The ability of sperms to donate electrons and cause spin label reduction is closely linked to their metabolic activity (11). It is thus possible to follow the kinetics

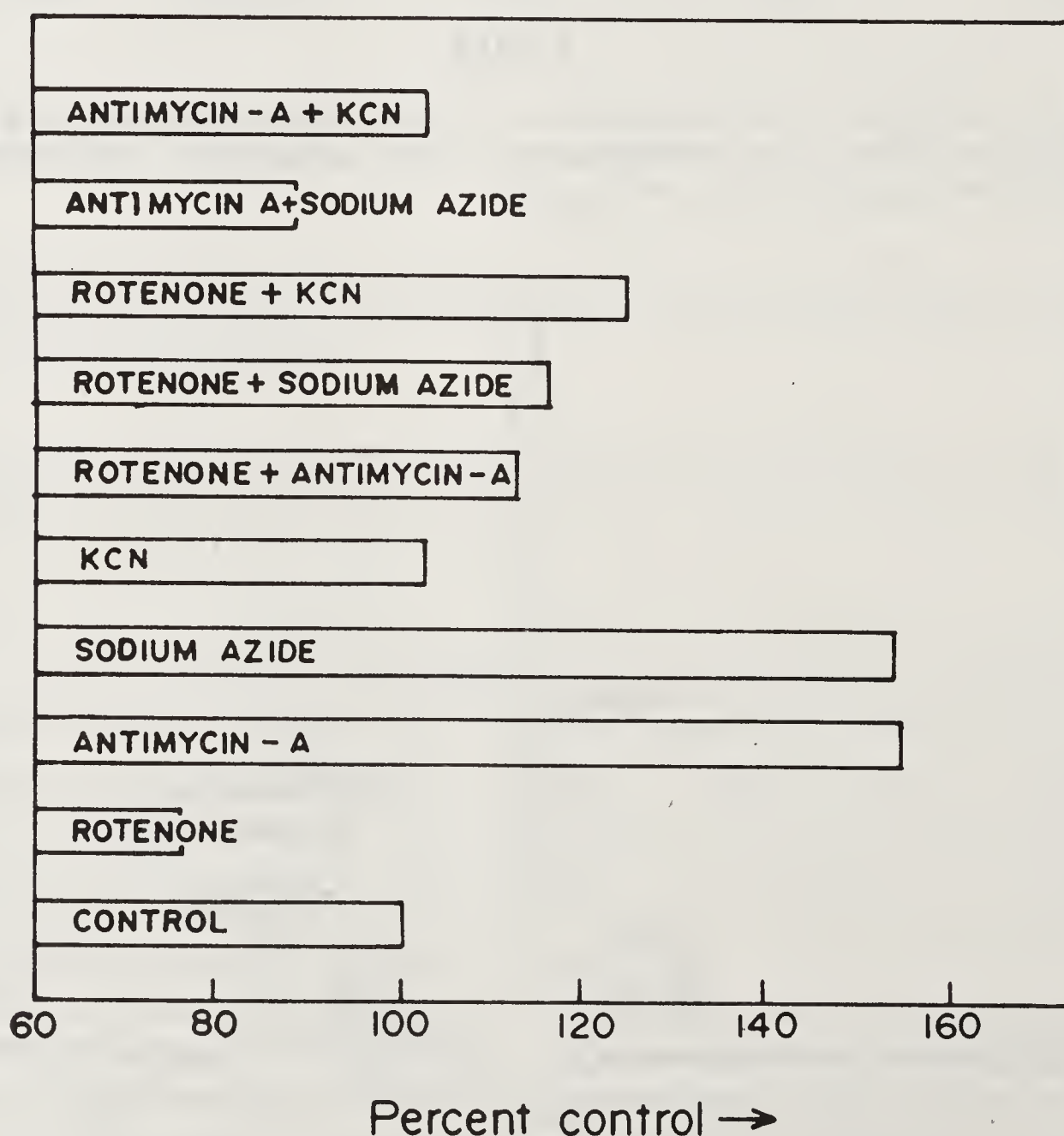


Figure 5. Effect of electron transport chain (ETC) inhibitors on the rate of 2,2,6,6-tetramethylpiperidine-N-oxyl (TEMPO) signal decay by human spermatozoa (10^9 sperms/ml) as compared to that of the control experiment expressed as a percentage

of spin label reduction by sperm cells and then study the factors influencing signal decay, which may provide useful information about the biology of sperms. It has been experimentally observed that the reduction of the spin label follows a pseudo first-order kinetics of the type $h(t) = h(0) e^{-kt}$ where $h(0)$ and $h(t)$ are signal heights at time 0 and t , and k is the rate constant (12, 13).

Figure 5 shows the influence of inhibitors of electron transport chain (ETC) on spin label reduction rate k . These inhibitors influence the ETC at different levels (14). Rotenone, antimycin-A and KCN and NaN_2 are known to cause blockage at flavoprotein (site I), cytochrome-b (site II) and cytochrome-a (site III), respectively. Rotenone ($2 \mu\text{g/ml}$) is observed to reduce the rate of reduction (k) to 73% of the control whereas, antimycin A ($100 \mu\text{l/ml}$) and NaN_2 (1 mM) increase the rate of reduction to about 150%. This indicates that spin label reduction occurs after site I and before site II, i.e. at ubiquinone level. Blockage of site I by rotenone restricts the influx of electrons in ubiquinone systems and causes lowering of rate of spin label reduction. Blockage at sites II and III prevents the passage of electrons to molecular oxygen leading to flooding of the ubiquinone site by electrons and thus resulting in spin label reduction. Combination of the two inhibitors gives an intermediate effect.

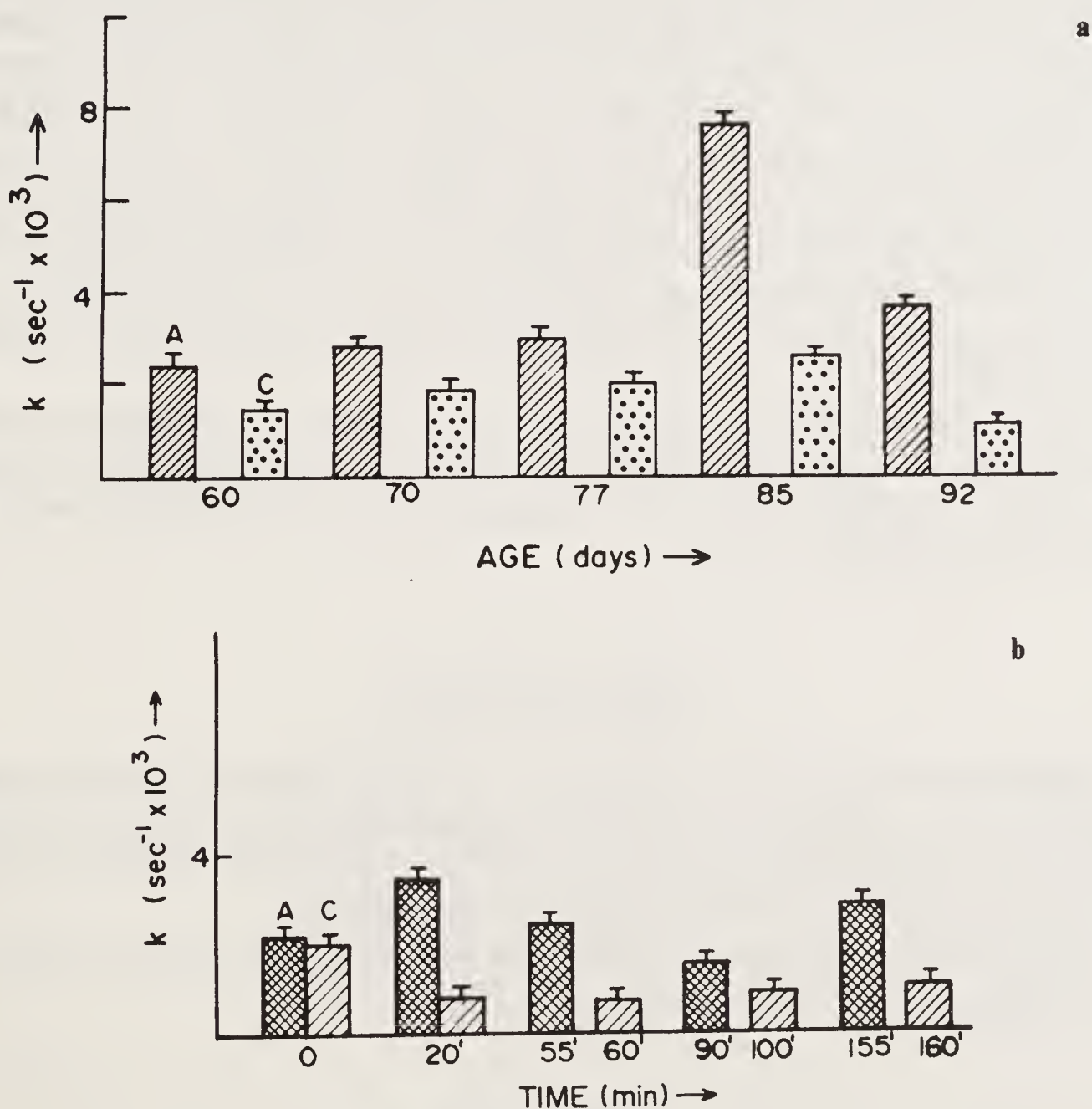


Figure 6. (a) The effect of sexual maturation on rate of TEMPO signal decay by rat spermatozoa from (A) caput and (C) cauda epididymis. Six rats were used for each experiment. (b) The effect of calcium (1.3 mM) on the rate of TEMPO signal decay by rat spermatozoa from (A) caput and (C) cauda epididymis. Cells were incubated with Ca^{+2} for different time intervals.

The effect of certain other factors, such as gossypol acetic acid (GAA, a known antifertility drug), anaesthetics and cold shock, on reduction kinetics has been studied for goat and human spermatozoa. GAA (100 μ M) increases the rate of reduction by 70% in human spermatozoa, the reason possibly being its damaging effect on plasma membrane and enhanced permeation of spin label. In contrast, it causes marginal increase (about 8%) in goat cells indicating its selective effect. Similar behaviour has been observed for anaesthetics (130% and 10% increase respectively) on the two types of systems.

Effect of cold shock on human spermatozoa shows no difference in reduction kinetics from that of control sample. This is consistent with the earlier findings that rapid cooling of semen by exposure to liquid nitrogen is a safe method for preservation of sample (15). However, in the case of goat spermatozoa the rate for spin label reduction is 3.5 times greater than that of control. This may be attributed to the increased accessibility of mitochondria to the spin label resulting from the alterations in the plasma membrane caused by cold shock.

Figure 6a shows the reduction kinetics of cells obtained from the caput and cauda epididymal regions of rat as a function of the animal's age. It is observed that rate of spin label reduction increases up to the age of 85 days, and thereafter declines slowly and levels off after 90 days. Figure 6b shows the result of a similar experiment as a function of Ca^{+2} concentration. Ca^{+2} is known to assist *in vitro* capacitation. The cells were incubated for different periods of time in the Ca^{+2} containing medium. It is observed that the k value drops almost to half of that of control at $t = 20$ minutes. The trend continues till $t \geq 60$ minute, after which it tends to increase though not much until 160 minutes. This pattern is parallel to Ca^{+2} uptake reported earlier (16), where in the first phase there is an accumulation of Ca^{+2} on the surface of sperm while in the second phase Ca^{+2} permeates through the membrane inducing the activation of sperms.

In short, it is possible to correlate the results obtained from magnetic resonance methods with the biology of the sperm cells and, thus, the technique can be developed into quality rating methodology and understanding of the biochemistry of this important class of cells.

Acknowledgements

The facilities provided by the 500 MHz FT-NMR National Facility, supported by the Department of Science and Technology, and located at T.I.F.R., Bombay, are gratefully acknowledged. We thank Mr. V. R. Bhagat for his help in ESR experiments. Sudha Srivastava wishes to thank CSIR for their support.

It is indeed a great pleasure to contribute this article in honour of Prof. G. N. Ramachandran, whom we all regard as our Guru.

References

1. Mann, T. & Mann, C. L. (1981) *Male Reproductive Function and Semen* (Springer-Verlag, Berlin, N.Y.).

2. Mann, T. (1977) in *Frontiers in Reproduction and Fertility Control*, eds. Greep, R. O. & Koblinsky, M. A. (MIT Press, Cambridge, Mass, and London) pp. 427–437.
3. Gadian, D. G. (1977) *Nuclear Magnetic Resonance and its Applications to Living Systems* (Oxford University Press, Oxford).
4. Chulavatnatol, M. & Haesungcharern, A. (1977) *J. Biol. Chem.* **252**, 8088–8095.
5. Moon, R. B. & Richards, J. H. (1973) *J. Biol. Chem.* **248**, 7276–7289.
6. Evans, R. W. & Setchell, B. P. (1979) *J. Reprod. Fertil.* **57**, 189–194.
7. Eng, L. A. & Oliphant, G. (1978) *Biol. Reprod.* **19**, 1083–1088.
8. Zaneveld, L. J. D. & Williams, W. L. (1970) *Biol. Reprod.* **2**, 363–368.
9. Bedford, J. M. (1972) in *Reproductive Biology*, eds. Balin, H. & Glasser, S. (Excerpta Medica, Amsterdam) pp. 338–345.
10. Chang, M. C. (1951) *Nature (London)* **168**, 697–699.
11. Chapman, D. A., Killian, G. J., Gelerinter, E. & Jarret, M. T. (1985) *Biol. Reprod.* **32**, 884–893.
12. Bahl, S., Bawa, S. R., Srivastava, S., Phadke, R. S. & Govil, G. (1988) *Physiol. Chem. Phys. Med. NMR* **20**, 183–188.
13. Fernandes, J. E., Hegde, U. C., Srivastava, S., Phadke, R. S. & Govil, G. (unpublished results).
14. Lehninger, A. L. (1982) *Principles of Biochemistry* (Worth Publishers, N.Y.) pp. 474 and 482.
15. Polge, C. (1979) in *Frozen Human Semen*, eds. Richardson, E., Joyce, D. & Symonds, M. (R. Coll. Obstet Gynaecol., London) p. 18.
16. Singh, J. P., Babcock, D. F. & Lardy, H. A. (1978) *Biochem. J.* **172**, 549–556.

Index of Contributors

- | | | | |
|---|-----|---|-----|
| Acharya A. Seetharama | | Coletti-Previero M. A. | |
| Polymerization of Hemoglobin S Amidated | | <i>see</i> Mattras H. | 385 |
| at the Contact Residue Glu 43(β) | 269 | Crisma Marco | |
| Anantharamaiah G. M. | | <i>see</i> Benedetti Ettore | 497 |
| <i>see</i> Segrest Jere P. | 597 | Cung Manh Thong | |
| Anantharamaiah G. M. | | 2D-NMR Conformational Analysis of | |
| <i>see</i> Venkatachalapathi Y. V. | 585 | Griselimycin, an Antituberculous Cyclo- | |
| Ashida Tamaichi | | depsipeptide | 611 |
| <i>see</i> Suzuki Atsuo | 171 | | |
| | | De Loof Hans | |
| Balaram P. | | <i>see</i> Segrest Jere P. | 597 |
| <i>see</i> Srinivasan N. | 59 | Demange Pascal | |
| Balasubramanian D. | | <i>see</i> Cung Manh Thong | 611 |
| <i>see</i> Luthra Manni | 281 | Dhingra Madan M. | |
| Bansal Manju | | <i>see</i> Rizo Josep | 469 |
| DNA Bending: A Natural Consequence of | | Dodson G. G. | |
| Base Sequence Dependent Variability | 347 | Conformational Changes at the B Chain N | |
| Benedetti Ettore | | Terminus in the Insulin Hexamer | 97 |
| The 3_{10} - and α -Helical Conformations in | | Dohlman Jan G. | |
| Peptides | 497 | <i>see</i> Venkatachalapathi Y. V. | 585 |
| <i>see</i> Toniolo Claudio | 511 | Doig Andrew J. | |
| Bhattacharyya Dhananjay | | Molecular Recognition in Peptides and | |
| <i>see</i> Bansal Manju | 347 | Proteins | 363 |
| Bianchi M. | | | |
| <i>see</i> Mattras H. | 385 | Easwaran K. R. K. | |
| Blackburn Warren D. | | Ionophores: Structure and Interaction in | |
| <i>see</i> Venkatachalapathi Y. V. | 585 | Relation to Transmembrane Ion-Trans- | |
| Blasio Benedetto Di | | port | 671 |
| <i>see</i> Benedetti Ettore | 497 | Efimov A. V. | |
| Boige grain R. A. | | Long-and Medium-Sized Irregular Regions | |
| <i>see</i> Mattras H. | 385 | in Proteins as Combinations of Small | |
| | | Standard Structures | 19 |
| Carey Robert I. | | Englander Joan J. | |
| Template Assembled Synthetic Proteins | | <i>see</i> Englander S. Walter | 245 |
| (TASP): A Chemical Approach to the | | Englander S. Walter | |
| Construction of New Proteins | 457 | Protein Interaction, Folding and Energetics | |
| Castro B. | | Studies by Hydrogen Exchange Labeling | 245 |
| <i>see</i> Mattras H. | 385 | | |
| Chait Brian T. | | Fehrentz J. A. | |
| <i>see</i> Acharya A. Seetharama | 269 | <i>see</i> Mattras H. | 385 |
| Chiche L. | | Fernandes Elvyra J. | |
| <i>see</i> Mattras H. | 385 | <i>see</i> Srivastava Sudha | 703 |
| Chothia Cyrus | | Frederick Christin A. | |
| Conformations for Strand Entry into | | <i>see</i> Williams Loren Dean | 295 |
| Parallel β -Sheets | 49 | | |
| Clark John D. | | Gessner Reinhard V. | |
| Helical Transitions in Peptides | 503 | <i>see</i> Williams Loren Dean | 295 |
| Cohen Fred E. | | Gierasch Lila M. | |
| <i>see</i> Troyer John M. | 31 | <i>see</i> Rizo Josep | 469 |

- Goodman M.
 see Venkatachalapathi Y. V. 523
- Govil Girjesh
 see Srivastava Sudha 703
- Gupta Kiran B.
 see Venkatachalapathi Y. V. 585
 see Segrest Jere P. 597
- Guptasarma P.
 see Luthra Manni 281
- Gutte Bernd
 The Design of Novel Proteins 447
- Hegde Umashashi C.
 see Srivastava Sudha 703
- Hiremath C. N.
 see Murthy M. R. N. 687
- Hodgkin Edward E.
 see Clark John D. 503
- Hol Wim G. J.
 Crystallographic Investigations of Glycolytic Enzymes from *Trypanosoma brucei*: Potential Starting Points for the Design of New Sleeping Sickness Drugs 215
- Hruby Victor J.
 The Ramachandran Plot and Beyond: Conformational and Topographical Considerations in the Design of Peptides and Proteins 429
- Hubbard R. E.
 see Dodson G. G. 97
- Imanishi Yukio
 Molecular Architecture by Using Cyclic Peptides for Ion-Transport Through Lipid Membrane 645
- Kannan K. K.
 Structural Perspectives in the Function of Carbonic Anhydrase 131
- Katakai R.
 see Venkatachalapathi Y. V. 523
- Khandke Lakshmi
 see Acharya A. Seetharama 269
- Kimura Shunsaku
 see Imanishi Yukio 645
- Klauser Stephan
 see Gutte Bernd 447
- Kodandapani R.
 see Vijayan M. 149
- Lesk Arthur M.
 see Chothia Cyrus 49
- Luthra Manni
 Crystallin Structure and Eye Lens Function 281
- Marraud Michel
 see Cung Manh Thong 611
- Marshall Garland R.
 see Clark John D. 503
- Mathews F. Scott
 Complex Flavoenzymes – Models for Electron Transfer Complexes 187
- Mattras H.
 Molecular Recognition in Aspartic Proteases 385
- Mayne Leland
 see Englander S. Walter 245
- McKinnie Russel E.
 see Englander S. Walter 245
- Mohan S.
 Stereochemical Studies on DNA and DNA–Drug Interactions 311
- Mohanty Debasisa
 see Bansal Manju 347
- Murthy M. R. N.
 Structure and Assembly of Viruses 687
- Mutter Manfred
 see Carey Robert I. 457
- Nagaraj R.
 Signal Sequences: The Relevance of Their Conformations and Membrane Perturbing Ability in the Sorting of Secretory Proteins 661
- Nandi Lilian C.
 see Singh Juswinder 1
- Nikiforovich Gregory V.
 see Hruby Victor J. 429
- Noble Martin E. M.
 see Hol Wim G. J. 215
- Paterson Yvonne
 see Englander S. Walter 245
- Pavone Vincenzo
 see Benedetti Ettore 497
- Pedone Carlo
 see Benedetti Ettore 497
- Phadke Ratna S.
 see Srivastava Sudha 703
- Ponnuswamy P. K.
 Solvent Accessibility of Nucleic Acids 335
- Poulos Thomas L.
 Structural Studies on the Mechanism of O₂ Peroxide Activation and Electron Transfer in Peroxidase and Cytochrome P-450 75
- Ramakrishnan C.
 see Srinivasan N. 59
- Rao M. Janardhan
 see Acharya A. Seetharama 269
- Rao V. S. R.
 Protein–Carbohydrate Interaction – A Theoretical Approach 411

- Read Randy J.
 see Hol Wim G. J. 215
- Rich Alexander
 see Williams Loren Dean 295
- Rizo Josep
 Peptide Models for Reverse Turns. The
 Role of Asparagine in the i Position of a β
 Turn 469
- Roberts G. C. K.
 Dynamic Aspects of Protein Specificity 401
- Santini Antonello
 see Benedetti Ettore 497
- Savithri H. S.
 see Murthy M. R. N. 687
- Schulz Georg E.
 Protein-Crystallography and Enzyme Cata-
 lysis 115
- Segrest Jere P.
 see Venkatachalapathi Y. V. 585
 Role of Basic Amino Acid Residues in the
 Amphipathic Helix: the Snorkel Hypothesis 597
- Singh Juswinder
 Towards Understanding How Side Chains
 Pack in Proteins 1
- Sivakama Sundari C.
 see Luthra Manni 281
- Sowdhamini R.
 see Srinivasan N. 59
- Srinivas Shamala K.
 see Venkatachalapathi Y. V. 585
 see Segrest Jere P. 597
- Srinivasan N.
 Analysis of Short Loops Connecting
 Secondary Structural Elements in Proteins 59
- Srivastava Sudha
 Metabolism and Maturation Studies of
 Intact Spermatozoa Using Magnetic Reso-
 nance Techniques 703
- Stroud Robert M.
 Structure and Function of Transmembrane
 Ion Channels 627
- Subramanian E.
 Peptides as Models for Protein-Folding 537
- Suzuki Atsuo
 Structure of Bowman-Birk Protease Inhi-
 bitor and its Complex with Trypsin 171
- Thornton Janet M.
 see Singh Juswinder 1
- Toniolo Claudio
 see Benedetti Ettore 497
 The Fully Extended Polypeptide Conforma-
 tion 511
- Troyer John M.
 Simplified Models for Understanding and
 Predicting Protein Structure 31
- Tsunogae Yasuo
 see Suzuki Atsuo 171
- Urry Dan W.
 Thermally Driven Self-Assembly, Molecular
 Structuring and Entropic Mechanisms in
 Elastomeric Polypeptides 555
- Vellieux Fred M. D.
 see Hol Wim G. J. 215
- Venkatachalapathi Y. V.
 A Guest-Host Approach to Oligodepsipep-
 tides: Studies of Oligoglutamates and
 Depsiglutamates in Trifluoroethanol 523
 Use of Synthetic Peptides in the Structural
 and functional Studies of Apolipoprotein
 A-I 585
 see Segrest Jere P. 597
- Verlinde Christophe L. M. J.
 see Hol Wim G. J. 215
- Vijayan M.
 Variability in Protein Hydration and its
 Structural Consequences – A Novel Appro-
 ach involving Water-Mediated Transforma-
 tions 149
- Vitoux Bernard
 see Cung Manh Thong 611
- Wierenga Rik K.
 see Hol Wim G. J. 215
- Williams Dudley H.
 see Doig Andrew J. 363
- Williams Loren Dean
 Molecular Structures of Spermine:Nucleic
 Acid Complexes 295
- Xiao B.
 see Dodson G. G. 97
- Yamane Takashi
 see Suzuki Atsuo 171
- Yathindra N.
 see Mohan S. 311

Subject Index

- AChR, 629–637, 640–642
 Nicotinic AChRs conduct, 628, 629, 633
 Actaplanin (A-4696), 367
 Adenylate kinases, 123, 127
 Adriamycin, 301–304
 α -Aminoisobutyric acid (Aib), 498, 511, 513, 514, 519, 546, 656
 Aib-like residues, 498
 α -Methylalanine (MeA, aminoisobutyric acid Aib), 503–506
 C $^{\alpha\alpha}$ -disubstituted α -aminoacids, 512, 513
 Alamethicin, 658
 Crystal structure of alamethicin, 504
 AMBER, 311, 319, 623
 Amino acids
 see under Ac $_n$ c, C $^{\alpha\alpha}$ -disubstituted α -amino acids, α -aminoisobutyric acid, unnatural amino acids
 1-Aminocycloalkane-1-carboxylic acid (Ac $_n$ c), 511–514, 519
 1-Amino naphthalene-4,6,8-trisulphonic acid (ANTS), 233
 Amphipathic helices, 585, 587, 591, 593, 597–606, 608, 667
 Angiotensin, 434, 437
 8-Anilinonaphthalene-1-sulfonate (ANS), 646, 649, 654, 655, 657
 Anthracyclines, 301, 304, 307, 308
 Antibiotics, 364, 365, 370, 381
 Antibiotic A35512, 367
 Penicillin, 364
 Peptaibol antibiotics, 504, 509
 Sulfonamides, 139, 140, 142
 Apolipoproteins, 591, 593, 594, 597, 604
 Apolipoprotein apo A-I, 585–587, 590, 591, 593, 594
 Disoidal complexes, 590, 593
 Disoidal particles, 588, 589, 591
 L-Arabinose binding protein (ABP), 412, 416
 Avoparcin, 367

 Bacteriophage MS2, 694, 696, 700
 Phage heads, 308
 Biological electron transfer, 77, 87, 92, 94
 Electron transfer complex, 187, 188
 Electron transfer mechanisms, 91
 Electron transfer proteins, 187
 Electron transfer reactions, 75
 Electron transport, 187

 Bradykinin, 433, 434, 436, 437
 Brownian dynamics, 90
 α -Bungarotoxin, 640, 641

 C $_5$ forms, 512, 515, 518
 C $_5$ conformation, 513–516, 519, 624
 C $_5$ structure, 511, 519
 Fully extended (C $_5$) conformation, 513
 Fully extended polypeptide conformation, 511, 515
 C $^{\alpha\alpha}$ -symmetrically disubstituted residues, 519
 C $^{\alpha\alpha}$ -disubstituted α -aminoacids, 512, 513
 C $^{\alpha\alpha}$ -symmetrically disubstituted side chains, 519
 C $^{\alpha\alpha}$ -symmetrically disubstituted α -aminoacids, 514
 Dialkyl amino acids, 504, 505, 507, 509
 Diphenylglycine (D ϕ g), 511, 513–518
 Dibenzylglycine (Db $_z$ g), 511, 513–519
 Diethylglycine (Deg), 511, 514–516, 518, 519
 α,α -disubstituted amino acids, 507
 Iva, 511, 513, 514, 519
 Calcimycin (A23187), 673, 676–683
 Camphor, 82
 Carbonic anhydrases, 131–144
 Zinc metalloenzyme, 131
 Catalase, 76, 77, 87
 CHARMM, 681
 Chromatin, 295, 308
 Chromophores, 282, 284, 286–288, 292
 Chymotrypsin, 117, 171, 172, 175, 176
 Circular dichroism CD, 447–450, 458, 462–464, 523, 526, 528, 531–534, 600, 601, 646, 657, 677–679, 682
 Collagen, 540, 546, 547
 Computational chemistry, 504
 Energy minimisation, 319–321, 323, 412, 414, 415, 418, 422, 424
 Concanavalin A (Con A), 418, 420, 421, 423
 Con A-ligand complexes, 422
 Conformations, 32, 38, 39, 41, 43, 49–54, 56, 57, 59, 64, 68–71, 470–473, 475–477, 483, 489–494
 β -Conformation, 662, 664
 Conformation of signal sequences, 662
 Conformation space, 34
 Conformational changes, 31

- Conformational energies, 513
Conformational energy calculations, 311, 312, 437, 511
Conformational families, 66, 68–70
Conformational fluctuations, 403
Conformational parameters, 499
Conformational space, 33, 40, 44
Conformational subsearches, 36
Conformational transition, 509
Conformations in peptides, 497
Ring flipping, 404
p-Cresol methylhydroxylase (PCMH), 187, 204, 206, 208–212
Crithidia fasciculata, 218, 221
Crystal structure analysis, 349, 357, 360
 Crystal structure of alamethicin, 504
 Crystal structures, 77, 83, 171, 297, 497
 Crystal structures of peptides, 497, 498, 501, 507
Crystallins, 281, 282, 284, 286–292
 Cataractogenesis, 281, 282, 286
 Cataractous, 284, 286, 287
 Eye lens, 281, 284, 286, 287, 289, 291, 292
 Light transmission, 281
Crystallographic studies, 80, 97, 134, 224, 230, 232, 234, 236, 239, 242, 402, 404, 449
 Crystal structures, 542, 544, 546, 547, 552, 675, 677
 Multiple isomorphous replacement method, 135
 Protein crystallography, 115, 116, 149, 177
 Protein crystals, 149, 152, 162, 167
 Superposition solvation map, 338, 339
Cyclic peptides, 240–242, 645–647, 649–651, 653, 656, 658
 Cyclic hexapeptides, 649, 651
 Cyclic octapeptides, 651, 652, 658
 Cyclo (L-Asp–L-Phe–L-Pro)₂, 649, 650
 Cyclo(L-Cys(Acm)–L-Phe–L-Pro)₂, 647, 648
 Cyclo (D-Leu–L-Glu–L-His)₂, 650
 Cyclo(L-Leu–L-Phe–L-Pro)₂, 647, 648
 Cyclo(L-Leu–L-Phe–L-Pro)₄, 647
 Cyclo(L-Lys(Z)–Sar)₄, 654, 655, 657
 Cyclo(L-Lyz(Z)–Sar–D-Leu–Sar)₂, 652, 653
 Cyclo(VPGVG)₃, 557, 560, 562, 565, 567–570
Cyclodepsipeptide, 611
Cytochrome b₅, 202, 203
Cytochrome c, 88, 90, 245–253, 354
Cytochrome c peroxidase (CCP), 75, 77–81, 87, 88, 90, 92, 93
Cytochrome P-450, 75–77, 81–84, 86, 87, 452
 P-450 cam, 85, 88
DAPI (4',6'-diamidino-2-phenyl indole), 326–332
DNA, 132, 137, 217, 295, 297, 301, 303–308, 311, 313, 318, 326, 330, 332, 335–339, 343–345
 A-DNA, 298, 300, 306, 307
 B-DNA, 307, 339
 Base pair stacking, 348
 Base pairs, 311, 313, 315, 318, 319, 321, 323, 325–327, 329, 332
 Base sequence effects, 348, 354, 356, 357
 Complementarity, 371, 372, 381
 Curvature, 348
 DNA:anthracycline complexes, 301–303, 306
 DNA base pair, 296
 DNA bending, 347, 348, 358
 DNA-binding, 449, 451, 651
 DNA binding domains, 406
 DNA binding proteins, 401, 450
 DNA–drug interactions, 311
 DNA structure, 311
 Kinetoplast DNA, 217
 Major groove, 313, 315, 318, 321, 323, 330
 Minor groove, 312, 315, 318, 323, 325, 328–332
 Nucleic acid, 335
 Nucleic acid complexes, 295
 Nucleic acid conformations, 295, 297
 Nucleic acid structure, 295, 306
 Nucleic acid structural motifs, 308
 Nucleotide, 122
 Propeller twists, 312, 313, 315, 316, 318, 319, 321, 323–326, 332
 Spermine:DNA complexes, 308
 Spermine:nucleic acid complexes, 297, 306
 Stacking energy, 348, 355, 356, 357
 Z-DNA, 304–308
Daunomycin, 301–303
De novo design, 447
 DDT-binding peptide, 453
 DDT-binding polypeptide, 452
 De novo design of proteins, 453, 457, 458, 464, 465
 Designed polypeptides, 449
 Protein design, 447
Depsipeptide, 523, 524, 526, 528–531, 534
 Cyclodepsipeptide, 611
 Decadepsipeptide, 532
 Oligodepsiglutamates, 529, 533
 Oligodepsipeptides, 523, 524, 528
Difluoromethylornithine, 219
Dihydrofolate reductase (DHFR), 401, 403
 E. coli DHFR, 402
 L. casei DHFR, 405, 407
 DHFR, 404, 406
Dipole moment, 506
Disulfide linkages, 176, 184, 185
 Disulfides, 119, 121, 122

- Disulfide bonds, 378, 380, 381, 449
- Drug design, 216, 219, 221, 222, 224, 242
- EPR/ESR spectroscopy, 80, 142, 704, 708
 - ENDOR spectroscopy, 80
 - Spin-labels, 703, 704, 708, 709
- Elastin, 555, 559, 565, 572, 573
 - Elasticity, 555, 565, 571, 572, 580
 - Elasticity in self-assembling, 562
 - Entropic elasticity, 555
 - Entropic elastomeric force, 556
- Electron density maps, 142, 220, 296, 297
- Entropy, 564, 571–577, 580
 - Entropic elasticity, 555
 - Rotational entropy, 371
 - Translational and rotational entropies, 370, 371, 381
- Enzyme models, 650
 - Cyclodextrin, 653, 654
- Enzyme catalysis, 115
 - Active site, 218, 224, 229, 231–236, 238, 242
 - Active triad, 390
- Electron microscopy, 600, 603, 604
- Enkephalins, 433, 436–438, 441, 442, 539
- Epitope, 246, 249, 250
- Flavo enzyme, 117
- Flavocytochrome b₂, 187, 194–204, 211, 212
- Fluorescence, 251, 386, 387, 600, 601, 606, 646, 649, 652, 655, 657, 665, 677, 679, 680, 682
 - Fluorogenic substrates, 385, 386, 388
- Glutathione (reduced) GSH, 121
 - Glutathione–spermidine, 218
- Glutathione (oxidised) GSSG, 117, 119, 121, 122
- Glutathione reductase, 117–121, 123, 128, 192–194, 211, 218
- Glyceraldehyde 3-phosphate dehydrogenase (GADPH), 223–226, 228
- Glycolate oxidase, 194, 203, 204, 211
- Glycolysis, 220
- Glycolytic enzymes, 215
- Glycoproteins, 219
 - Variable surface glycoprotein (VSG), 219, 220
- Glycosomal enzymes, 222, 224, 225, 227, 228, 230, 236
 - Glycosomal TIM, 235
- Glycosomes, 218, 220–224
- Gramicidin S, 648
- Griselimycin, 611–618, 621, 623
- Guanylate kinase, 125, 126
- Hard-sphere model, 31
- Hemoglobin (Hb), 245, 246, 258–263, 266, 267, 277
 - Sickle cell hemoglobin (Hbs), 269, 270–272, 274, 276, 278
- Helichrome, 655, 656
- Helix/helices
 - 3₁₀/α-Helical structure, 505, 513, 519, 523, 530, 534
 - 3.6₁₀-Helix, 503
 - 3₁₀-Helices, 497, 499–506, 509, 546
 - 3₁₀- and α-helical types, 498
 - Amphipathic helix, 585, 587, 591, 593, 597–599, 604, 606, 667
 - Averaged helical parameters, 501
 - Crystal structures of helical peptides, 507
 - α-Helical conformation, 662, 663, 664
 - Helical transitions, 503
 - α-Helices, 59, 497, 499–506, 509, 546, 547
 - α-Helix bundle structures, 41
 - Helix dipole, 606, 608
 - Helix geometry, 497
- Heme, 75, 187, 188, 195, 197, 205, 209
- Heme proteins, 75
- HIV protease, 385, 388
 - HIV-1 protease, 394, 395
- Hydrogen bonds, 1, 6, 7, 9, 12, 13, 49–52, 137, 138, 142, 149, 153–157, 163–165, 175–177, 181, 183, 185, 252, 256, 257, 295, 298, 300, 301, 302, 304, 306, 315, 316, 318, 319, 321, 323, 326–332, 412, 414, 418–423, 469–480, 483–485, 487–489, 491–494, 503, 506, 511, 514–516, 518, 519, 523, 528, 532, 540, 541, 544, 545, 547–549, 616, 623, 624, 673–675, 679, 682
 - H-D exchange, 245, 246, 249, 255, 256, 267
 - H-exchange, 247, 248, 251, 252, 254, 257, 259, 261, 264, 265
- Hydropathic index, 665, 666
- Hydrophobic effect, 373, 374, 376, 380, 381
- Infrared spectroscopy (IR), 141, 142, 513, 518, 623, 649, 678
- Inhibitors, 172
 - AB-I inhibitor, 173
 - Active site inhibitors, 238
 - Bowman-Birk type inhibitors, 171, 174, 183
 - Pepstatin, 385, 392
 - Trypsin AB-1 complex, 173, 175, 178, 180
- Insulin
 - Insulin hexamer, 97–99, 101, 108, 111
- Inverse temperature transition, 555, 560, 563, 564, 568, 569, 572
- Ion channels, 627–629, 631–637, 640, 642, 645, 658

- Channel blockers, 632
- Channel conductance, 631, 633, 634
- Channel protein, 464, 465
- Chloride conducting channels, 637
- Conductance, 630, 632, 634–636
- Glutamate receptors, 637
- Ion transport, 671, 672
- Ionophore, 645, 646, 651, 671–673, 676, 678–681, 683
 - A23187, 676–683
 - Valinomycin, 672–677, 683
- Iron porphyrin IX
 - Iron protoporphyrin IX, 75
- Isoenzymes, 132
 - Isozymes, 134
- Iron–sulphur proteins
 - Fe₄S₄ cluster, 191
 - (4Fe-4S) cluster, 187, 189, 190, 192
 - (4Fe-4S) center, 188
- Karplus equation, 619, 620
- Kinases, 122, 123
 - Adenylate kinase, 123, 127
 - Guanylate kinase, 125, 126
 - Nucleotide kinases, 124, 126, 128
- Knowledge-based modelling, 36, 37
- Lecithin:cholesterol acyl transferase (LCAT), 585, 591
 - LCAT activating properties, 587
 - LCAT activation, 590, 593–595
- Leishmania*, 221
 - Leishmania mexicana*, 218
- Leitstruktur, 118, 119
- Lipid bilayer membranes, 665
 - Lipid bilayer, 665, 667
 - Lipid membrane, 645
- Lysozyme, 424, 425, 426
 - Hen egg white lysozyme, 153, 424
 - Lysozyme saccharide complexes, 425
 - Monoclinic lysozyme, 167, 168
 - Tetragonal lysozyme, 153, 164, 165, 167
 - T4 lysozyme, 425
 - T4 lysozyme saccharide complexes, 425
- Marcus theory, 91, 92
- Mass spectrometry, 271
 - Mass spectral analysis, 272
- Melanotropin, 437
 - α-Melanotropin, 435, 436
- Membranes, 628–632, 635, 636, 642, 661, 666, 672, 673, 675, 684
 - Biological membranes, 671
 - Membrane proteins, 627
 - Phospholipid membrane, 646, 648, 649, 652, 658
- Molecular dynamics, 32–34, 39, 44, 233, 425, 436, 440, 472, 476, 477, 479, 480, 482, 486–493, 577, 579, 623
 - AMBER, 311, 319, 623
 - CHARMm, 681
 - Molecular dynamics simulations, 426, 509
- Molecular mechanics, 311, 319, 321, 395, 440, 442, 575
 - Minimized energy, 681
 - Molecular mechanics potential energy, 39
- Molecular modelling, 394, 412
 - Computer modelling, 411, 412, 414, 418, 423, 425
 - Computer simulation, 411, 412, 425
 - Modelling, 38
 - Modelling of water, 35
- Molecular recognition, 363–365, 370, 371, 375, 381, 385, 401, 412
 - Induced fit, 404, 412
 - Methotrexate, 402, 405–406
 - T4 lysozyme–saccharide complexes, 425
- Monte Carlo algorithm, 42
- Multiple isomorphous replacement method, 135
 - MIR phases, 142
- Mutagenesis, 79, 80, 84, 92, 93
 - Site directed mutagenesis, 85, 132, 141, 143, 145, 402, 407, 645
- NMR spectroscopy (NMR), 49, 141, 142, 144, 246, 247, 252, 254, 255, 257, 401, 402, 407, 437, 439–441, 472, 474–476, 478–483, 486–499, 491–493, 506, 523–526, 528, 532, 534, 538, 611, 612, 623, 676–679, 682, 703
 - ¹H NMR, 513, 518
 - ¹³C NMR, 706, 707
 - ³¹P NMR, 704–707
 - 2D NMR, 452
 - COSY, 611, 612
 - HOHAHA, 611
 - H-exchange, 247, 248, 251, 252, 254, 257, 259, 261, 265
 - H–D exchange, 246, 249
 - Intermolecular nuclear Overhauser effects (NOEs), 40
 - Interproton distance, 472, 475, 479, 480, 482, 486–489, 492, 493
 - Isotope-editing experiments, 403
- Nuclear Overhauser effects (NOE), 403, 405–407, 465, 474, 475, 479, 480, 482–487, 492, 493, 611–613, 620, 623, 624, 680
- Neurotransmitters, 429–431, 435, 438
- Neutrophil(s)
 - Neutrophil Fc-receptor mediated activation, 590

- Neutrophil activation, 587, 589, 595
- Neutrophil regulation, 595
- Oxidative phosphorylation, 77
- Oxygen activation, 81
 - O₂/H₂O₂ activation, 75
 - Oxygen and peroxide activation, 77
 - Peroxide activation, 87
- Oxytocin, 435
- Parasite enzymes, 216
- D-Penicillamine (D-Pen), 437
 - Penicillamine ($\beta\beta$ -dimethylcysteine), 429
- Penicillin, 364
- Pepsin, 385, 387–398
- Pepsinogen, 394
- Pepstatin, 385, 392
- Peptaibol antibiotics, 504, 509
- Peptides
 - Amphipathic peptides, 589, 593
 - Crystal structures, 497, 498, 501, 507
 - Cyclic peptides, 645–647
 - Cyclodepsipeptides, 611
 - Glutamate oligopeptides, 526, 527
 - Oligo- γ -alkyl glutamates, 532, 534
 - Oligoglutamate analogs, 523
 - Peptide antibiotics, 648
 - Peptide-DMPC particles, 589
 - Peptide:DMPC complexes, 588
 - Peptide:egg lecithin complexes, 590
 - Peptide hormones, 429–431, 435, 438
 - Peptide-peptide binding, 364, 370, 375, 376
 - β -Structures, 532
- Peptidomimetics, 437, 438
 - Peptidomimetic design, 436
- Peroxidases, 76, 77, 79, 80, 83, 84, 87
- Phospholipase A₂ (PLA₂), 401, 450
- Photosensitization, 289, 291, 292
 - Photosensitizers, 286–289
- Polyamines, 218, 219, 295, 297, 306, 308, 695
 - Glutathione-spermidine, 218
 - Spermidine, 219
 - Spermine, 295, 297–308
 - Spermine:DNA complexes, 308
 - Spermine:nucleic acid complexes, 297, 306
 - Spermine:phosphate, 298
 - Spermine:yeast tRNA, 306
- Polymerization, 269, 278, 279
 - Polymerization reaction, 270
- Polypeptides
 - Poly(Lys-Aib-Leu-Aib), 656, 658
 - Poly(VPGVG), 555, 557, 560–566, 569–577, 580
 - Polyglycine II, 546, 547
 - Polypeptide and polydeoxyribonucleotide synthesis, 447
 - Polypeptide conformation, 511, 519
 - Polypeptide structures, 37
 - Polyproline II, 546
 - Oligo- γ -alkyl glutamates, 532, 534
 - Oligodepsiglutamates, 529, 533
 - Oligoglutamate analogs, 523
 - Random coil structures, 534
 - β -Spiral, 561, 562, 564, 565, 569–571, 573, 575–577, 580
- Prediction of protein structure, 32, 36, 37, 44
 - Lattice models, 40
- Proteases, 389
 - Aspartic protease, 385, 386, 388, 394
 - Chymotrypsin, 117, 171, 172, 175, 176
 - HIV protease, 385, 388, 394, 395
 - New protein molecules, 458
 - Pepsin, 385, 387–398
 - Retroviral aspartic proteases, 398
 - Retroviral proteases, 394, 396
 - Serine proteases, 184
 - Trypsin, 171, 172, 175, 176, 178–184
- Proteins
 - (4Fe-4S) centre, 188
 - Allosteric, 258, 262, 263, 265, 363
 - Allosteric activity, 266
 - Artificial protein, 447, 462, 464
 - Artificial repressors, 449
 - Concanavalin A (conA), 418
 - Dynamic processes in proteins, 407
 - Folding, 245, 252, 255
 - Glycine-rich loop, 125
 - $\alpha\alpha$ Hairpins, 19
 - $\beta\beta$ Hairpins, 19, 22–24, 28
 - Hydration of proteins, 149
 - Hydration shell, 165, 167, 168, 345
 - Irregular regions, 19, 20, 24, 25, 28, 29
 - Knowledge-based approach to modelling, 37
 - Linker conformation, 66
 - Linkers, 63–65, 67–71
 - Loops, 19, 20, 59, 63, 64, 66, 68, 70, 71
 - Membrane proteins, 627
 - Packing density, 37
 - Packing preferences, 6, 13, 17
 - Pea lectin, 423
 - Protein crystallography, 115, 149
 - Protein crystals, 149, 152, 162, 167
 - Protein *de novo* design, 457
 - Protein design, 447
 - Protein dynamics, 37
 - Protein folding, 245, 250, 252, 254, 364, 365, 376, 378, 438, 447, 458, 460, 537
 - Protein hydration, 150, 166
 - Protein modelling, 32

- Protein specificity, 401
- Protein-carbohydrate complexes, 425
- Protein-carbohydrate interaction, 411, 414, 424
- Protein-crystallography, 116, 117
- Protein-ligand interactions, 411, 412, 416
- Protein-monoclonal antibody interaction, 267
- Protein-protein interactions, 246
- Secondary structure, 59, 61–64, 66, 67, 71, 385
- Secondary structure elements, 137
- Sequence dependent variability, 347
- β -Sheet structures, 49, 50, 52, 53, 55, 56, 532
- Small standard structures, 19–24, 26–28
- Strand, 59–51, 53, 54, 56, 57, 61, 62, 68, 69
- β -Strands, 59
- Strand entry, 49
- Targetting of proteins, 661, 668
- Translocation of proteins, 661
- Water bridges, 158, 159
- Proton shuttle, 143, 144
- Putidaredoxin, 83, 88

- Quasi-equivalence, 689, 696
- Quasi-equivalent bonding, 691

- RNA, 217, 336–339
 - RNA binding protein, 449
 - Spermine:yeast tRNA, 306
 - tRNA, 306, 335
 - tRNA^{phe}, 340–342
 - Yeast-tRNA^{phe}, 297–209, 306
- Ramachandran plot, 20, 52–54, 100, 177, 429–438, 443, 627, 628
- Contact criteria, 414, 419
- Ramachandran (ϕ, ψ) space, 138
- Ramachandran diagram, 97, 102, 401, 448, 539, 540
- Steric map, 66, 415, 418–421
- Raman spectroscopy, 75
- Retina, 281, 282, 284, 289, 291
- Cataractogenesis, 281, 282
- Refractive index, 281
- Refractivity gradient, 282
- Reverse turns, 21, 28, 548, 549
 - Asx turn, 478, 484, 485, 489
 - C₇(γ) turn, 528
 - γ -Turn, 469, 474, 475, 485
 - β -Turns, 469–479, 483–486, 488, 489, 491–494, 560–562, 565, 568, 571, 572, 574, 575, 579
- Ribonuclease A/(RNaseA), 448
- 63-Residue analogue of RNase A, 448

- Ristocetin, 365, 367, 369, 370, 375, 376, 381
- Ristocetin A, 367, 368, 369, 371

- Secondary metabolites, 364
- Self-assembly, 555, 556, 559, 560, 572
 - Elasticity in self-assembling, 562
 - Thermally driven self-assembly, 555, 556, 565
- Serine proteases, 171, 184
 - Chymotrypsin, 117, 171, 172, 175, 176
 - Trypsin, 171, 172, 175, 176, 178, 182–184
- Side chain-side chain interactions, 5, 157
 - Side chain packing, 1, 3, 17
 - Side chain-atom interactions, 9
 - Side chain-side chain pair types, 4
 - Side chain interactions, 105
- Signal sequences, 661–668
 - Conformation of signal sequences, 662
- Snorkel hypothesis, 597
 - Snorkel, 607, 608
 - Snorkel model, 604
- Solvent accessibility, 335, 337, 338, 340, 341
 - Solvent accessible areas, 342
- Somatostatin, 435, 437, 438, 440
- Spermatozoa, 703–705, 707, 709, 710
 - Capacitation, 705, 710
 - Sperm motility, 703
- Super-secondary structural motif, 59, 71
 - 4 Helix bundle, 461, 462, 464, 655
 - 8-Stranded antiparallel beta-barrel, 692, 700
 - Anti-parallel β structures, 541, 544, 545
 - $\beta\beta$ -Arc loops, 22, 28
 - $\beta_8\alpha_8$ -Barrel, 190, 191, 194, 198, 199, 203, 211
 - β Barrel, 41
 - Barrels, 53, 54
 - Glycine-rich loop, 125
 - Helix bundle structures, 41
 - $\beta_8\alpha_8$ -Motif, 203
 - Parallel β -barrel, 229
 - Parallel β structure, 540, 542
 - Parallel β -sheets, 49–57
 - Super-secondary structure, 62
- Substance P, 434

- Template assembled synthetic
 - proteins (TASPs), 457–459, 461, 462, 464, 465
 - TASP molecules, 460, 462, 466
 - Template-assembled proteins, 645
- Tetragastrin, 434
- Thyrotropin releasing hormone, 438
- Transition state analogue, 235, 237
- Trifluoroethanol (TFE), 523, 525–529, 531–533

- Trimethoprim, 403, 404, 405
 Trimethylamine dehydrogenase (TMADH), 187–195, 203, 211, 212
 Triose phosphate isomerase (TIM), 223, 228–234, 236–241
 Triose phosphate isomerase active site, 239
 Trp repressor, 401, 406
Trypanasoma brucei (*T. brucei*), 215, 219, 221, 222, 224, 225, 227–236, 238–241
 Trypanosome, 217–222, 229, 233, 239, 242
 Trypanosomal TIM, 227, 230, 234, 238, 240, 242
 Trypanosomal glycolysis, 222
 Trypanosome *Crithidia fasciculata*, 233
 Trypanothione, 218, 219
 Trypanothione reductase, 218
 Trypsin, 171, 172, 175, 176, 178, 182–184
 Trypsin AB-1 complex, 173, 175, 178, 180
 Trypsin AB-I interactions, 182
 Tryptic peptide mapping, 271
 Tryptic digest, 272
 Tryptic map, 276
 Tryptic peptides, 273, 277
 1,2,3,4-tetrahydroisoquinoline-2-carboxylic acid (Tic), 429, 439, 440, 442
 Tuftsin, 436
- UV difference spectroscopy, 447
 Unnatural amino acids, 457, 460
 1-Aminocycloalkane-1-carboxylic acids, 511, 513, 514, 519
 $\alpha\alpha$, Dialkyl amino acids, 504, 505, 507, 509, 514–516
 Diethylglycine (Deg), 511, 514–516
 Dipropylglycine (Dpg), 511, 513–516, 518, 519
 Diphenylglycine (D ϕ g), 511, 513, 514, 518
 Isovaline (Iva), 511, 513, 514, 519
- Valinomycin, 646, 647, 670–677, 683
 Van der Waals surface, 335, 336, 342, 345
 Vancomycin, 364–368, 375, 376, 381
 Viruses
 Animal and insect viruses, 695
 Bacteriophage m82, 694
 Bean pod mottle virus (BPMV), 693, 694
 Belladonna mottle virus (BDMV), 694, 695
 Black beetle virus (BBV), 696
 Brome mosaic virus (BMV), 697, 699
 Carlaviruses, 691
 Cowpea chlorotic mottle virus (CCMV), 697
 Cowpea mosaic virus (CPMV), 693, 699
 Foot and mouth disease virus (FMDV), 695
 Humans rhinovirus, 695
 Icosahedral asymmetric unit, 693
 Icosahedral symmetry, 688, 689
 Icosahedral viruses, 700
 Picornaviruses, 695, 698, 699
 Plant viruses, 687
 Polio virus, 695
 Polyoma virus, 696, 698
 Potato virus X (PVX), 691
 Potato virus Y (PVY), 691
 Rod-shaped viruses, 691
 Satellite tobacco necrosis virus (STNV), 693, 694
 Small RNA viruses, 687
 Small ssRNA containing viruses, 687
 Southern bean mosaic virus (SBMV), 689, 693, 694, 696, 698
 Spherical plant viruses, 692
 Tobacco mosaic virus (TMV) 688, 690, 691, 695, 697, 698, 700
 Tomato bushy stunt virus (TBSV), 692–698
 Turnip crinkle virus (TCV), 693
 Turnip yellow mosaic virus (TYMV), 694, 695
 Virus assembly, 697
 Virus evolution, 699
 X-ray studies on viruses, 690
- Water-mediated transformation, 150–152, 162, 164–166, 168
 Water structure, 149, 157
- X-ray crystallography, 215, 222, 297, 401, 407
 X-ray crystal structure, 75, 79, 84, 87, 120, 298, 306, 308, 327, 551
 X-ray crystallographic studies, 411, 414, 426, 537, 538
 X-ray diffraction, 116, 128, 142, 145, 411, 422, 425, 513, 514, 519, 676, 687
 X-ray structure analysis, 115
 X-ray techniques, 412

stereochemical studies on DNA; molecular recognition in peptides and proteins; peptide and protein design; peptide conformations in model systems; elastomeric polypeptides; apolipoprotein models and amphipathic helices; structure and function of transmembrane ion channels; ion transport; signal peptides; and virus structures and assembly. This coverage provides a broad survey of contemporary research in the area of biomolecular structure and interactions.

FRONT COVER Model of the nicotinic acetylcholine receptor-family ion channel, showing the inferred quasisymmetric pentameric arrangement of the M2 α -helical domains of receptor subunits. (Picture courtesy Robert M. Stroud, University of California, San Francisco)

BACK COVER A Ramachandran ϕ - ψ map. A pair of linked peptide units defining the torsion angles ϕ and ψ is superimposed. (Picture courtesy C. Ramakrishnan and K. P. Viji, Indian Institute of Science, Bangalore)

